2005-03

Accuracy of western North Pacific tropical cyclone intensity guidance

Blackerby, Jason S.

Monterey, California. Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

ACCURACY OF WESTERN NORTH PACIFIC TROPICAL CYCLONE INTENSITY GUIDANCE

by

Jason S. Blackerby

March 2005

Thesis Advisor:	Russell L. Elsberry
Second Reader:	Mark A. Boothe

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** March 2005	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE: Accuracy of Western North Pacific Tropical Cyclone Intensity Guidance		**5. FUNDING NUMBERS**	
6. AUTHOR(S) Blackerby, Jason S.			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000		**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A		**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited		**12b. DISTRIBUTION CODE** A	

13. ABSTRACT (maximum 200 words)

Consensus methods require that the techniques have no bias and have skill. The accuracy of six statistical and dynamical model tropical cyclone intensity guidance techniques was examined for western North Pacific tropical cyclones during the 2003 and 2004 seasons using the climatology and persistence technique called ST5D as a measure of skill. A framework of three phases: (i) initial intensification; (ii) maximum intensity with possible decay/reintensification cycles; and (iii) decay was used to examine the skill.

During both the formation and intensification stages, only about 60% of the 24-36 h forecasts were within +/- 10 kt, and the predominant tendency was to under-forecast the intensity. None of the guidance techniques predicted rapid intensification well. All of the techniques tended to under-forecast maximum intensity and miss decay/reintensification cycles. A few of the techniques provided useful guidance on the magnitude of the decay, although the timing of the decay was often missed. Whereas about 60-70% of the 12-h to 72-h forecasts by the various techniques during the decay phase were within +/- 10 kt, the strong bias was to not decay the cyclone rapidly enough. In general the techniques predict too narrow a range of intensity changes for both intensification and decay.

14. SUBJECT TERMS Tropical Meteorology; Tropical Cyclone Intensity; Tropical Cyclone Intensity Change Forecast Techniques; Numerical and Statistical Intensity Change Guidance	**15. NUMBER OF PAGES** 127
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

ACCURACY OF WESTERN NORTH PACIFIC TROPICAL CYCLONE INTENSITY GUIDANCE

Jason S. Blackerby
Captain, United States Air Force
B.S., Penn State University, 1995

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
March 2005

Author: Jason S. Blackerby

Approved by: Russell L. Elsberry
Thesis Advisor

Mark A. Boothe
Second Reader

Philip A. Durkee
Chairman, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Consensus methods require that the techniques have no bias and have skill. The accuracy of six statistical and dynamical model tropical cyclone intensity guidance techniques was examined for western North Pacific tropical cyclones during the 2003 and 2004 seasons using the climatology and persistence technique called ST5D as a measure of skill. A framework of three phases: (i) initial intensification; (ii) maximum intensity with possible decay/reintensification cycles; and (iii) decay was used to examine the skill.

During both the formation and intensification stages, only about 60% of the 24-36 h forecasts were within +/- 10 kt, and the predominant tendency was to under-forecast the intensity. None of the guidance techniques predicted rapid intensification well. All of the techniques tended to under-forecast maximum intensity and miss decay/reintensification cycles. A few of the techniques provided useful guidance on the magnitude of the decay, although the timing of the decay was often missed. Whereas about 60-70% of the 12-h to 72-h forecasts by the various techniques during the decay phase were within +/- 10 kt, the strong bias was to not decay the cyclone rapidly enough. In general the techniques predict too narrow a range of intensity changes for both intensification and decay.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First and foremost, I must offer my appreciation to my wife, Lisa, who accepted the long hours bringing the thesis to a close; without her support, the thesis process would have been much more burdensome. Of course, Professor Elsberry brought the research into focus, and eagerly contributed his thoughts to the project; I certainly benefited from his critical eye and patience during the editing process. Mark Boothe served as a fantastic research assistant and an excellent sounding board; his computer expertise yielded many of the analysis tools used in the thesis. LCDR Dave Brown took personal time to review the thesis and offer improvements. Capt Kevin LaCroix at AFWA/DNXM answered some key questions regarding the Air Force MM5 model. Finally, all of my family and friends deserve recognition for their support through conversation and prayer.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. MOTIVATION

1. Intensity Guidance

Tropical cyclone (TC) intensity forecasts have not improved commensurate with the increasing accuracy of TC track forecasts. While recent research efforts and tropical operational forecasting practices have increased the accuracy of three-, four-, and five-day TC track forecasts, contemporary TC intensity forecasts have a large inherent uncertainty (DeMaria and Kaplan 1999). Even though improvements are planned for the various numerical (higher resolution, improved physics, and faster computing times) and statistical (larger samples and additional predictors) models, shorter-term 'corrective' efforts such as better utilization of this guidance must also be considered. For example, Carr and Elsberry (1994) demonstrated that understanding model TC track errors could lead to more accurate forecasting, which prompted the development of the Systematic Approach Forecast Aid (SAFA). Once applied at the Joint Typhoon Warning Center (JTWC), SAFA became a major contributor to improved official track forecasts between 1999 and 2002 (Jeffries and Fukada webpage reference, cited 2005).

Unfortunately, operational forecasters have not been able to realize the same success with TC intensity forecasting, where large intensity errors become evident beyond 24 h (Table 1). Although some small year-to-year variability about the five-year means is noted for the 24-, 48-, and 72-h intensity forecasts, this variability is much larger for the 96- and 120-h forecasts, which have been officially released only since May 2003. The tendency for the longer-term forecasts errors to level off is due to at least two factors: (i) a decreasing number of forecasts verified, and only for those 'well-behaved' storms; and (ii) forecasts of storms that undergo extratropical transition and continue over land are not verified.

JTWC Absolute Mean Intensity Error (kt)					
Year	24h	48h	72h	96h	120h
2004	11	17	20	22	26
2003	11	16	19	21	19
2002	10	17	22	32	37
2001	11	17	21	29	28
2000	12	19	24	24	30
Five-year Mean	11	17	21	26	28

Table 1.1 Mean absolute intensity errors (kt) at the JTWC. The table was adapted from the JTWC forecast statistics webpage. [Available online at http://www.npmoc.navy.mil/jtwc/climostats/fcsttrkerr.html (Current as of 18 Feb 05)].

Typhoon (hurricane) intensity prediction is arguably more difficult than track prediction alone, and the two forecasts are interrelated since the current storm location and a future path will determine the atmosphere/ocean environmental forcing on the storm. That is, a track into a favorable environment most likely leads to intensification, while a hostile environment ultimately leads to decay. Thus, an accurate intensity forecast requires an accurate track forecast. Internal processes such as eyewall replacement cycles and external interactions with mid-latitude systems or the ocean also play an important role in the intensity changes of TCs.

Figure 1.1 is an illustration of intensity guidance errors produced by one intensity model for a specific storm during 2003. This example clearly illustrates that skillful (reliable) TC intensity guidance may not be available from one forecast interval to the next. As will be described below, this example illustrates several types of intensity errors. The question is whether the technique consistently makes similar errors in the same stage of that life cycle. That is, can the tropical cyclone forecaster be given guidance as to when (and when not) the techniques will be erroneous? If yes, then the selective consensus approach of Carr and Elsberry that contributed to dramatic improvements in track forecasting might be adaptable for intensity forecasting. That is, consensus intensity forecasts may be generated as new intensity guidance becomes available in conjunction with a well-developed meteorological knowledge base regarding storm structure/motion, known technique biases/errors, and forecaster judgments.

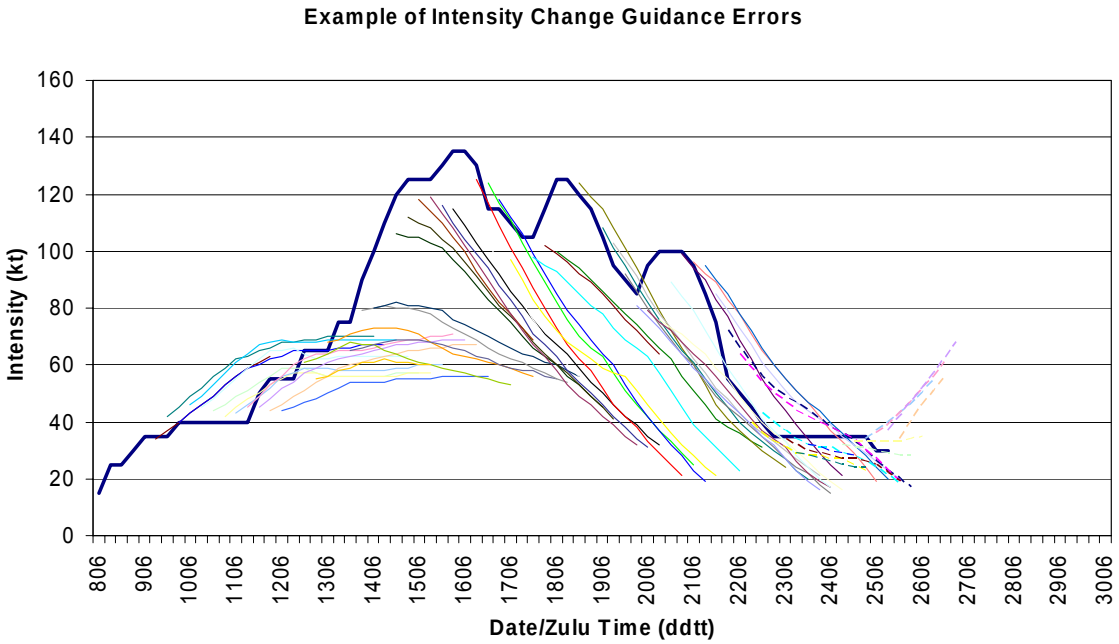

Figure 1.1 Statistical Typhoon Intensity Prediction Scheme (STIP) guidance for Storm #2 during 2003. The heavy blue line represents the observed (best track) storm intensity. Colored lines illustrate intensity forecasts each 6 h. The errors may be described as: early over-intensification; subsequent under-intensification; miss 'rapid' intensification; miss peak intensity; miss decay/reintensification cycles; premature decay; miss 'rapid' decay; and incorrect timing and rate of decay.

2. Ramifications of TC Intensity Forecasts

Tropical meteorology and oceanography (METOC) considerations are myriad in the western North Pacific Theater of Operations. While there are significant military aviation hazards associated with tropical systems, the preeminent threat to life and property is at sea. Due to the sheer number of vessels and logistical complexities, the United States Navy arguably remains most vulnerable to TC activity—especially in the vast Pacific Ocean. Moreover, fixed military installations often lie in the path, and most certainly come under the influence, of tropical cyclones. Some of the most prominent examples are: Kadena Air Base, Okinawa; Yokota Air Base, Japan; Yungson Air Base, Republic of Korea; Yokosuka Naval Installation, Japan; and Sasabo Naval Installation, Japan.

Each TC forecast initiates a series of decision-making processes through varying levels of military and civil organizations. Planners must subsequently modify shipping routes, curtail military maneuvers, or cancel commercial fishing operations to account for tropical weather impacts: dangerous significant wave heights; extremely high near-surface wind speeds; and near-zero visibility in heavy rain. Inaccurate TC track and intensity forecasts stress the socio-economic decision-making process, and lead to significant costs for evacuation (false alarm) or even greater losses due to damaged materiel—even loss of life—if TC effects are not well forecast.

B. TROPICAL CYCLONE INTENSITY FORECAST EXAMPLES

1. Hurricane Charley (Florida, U.S., 2004)

Hurricane Charley intensified rapidly prior to landfall on the west coast of Florida on 13 August 2004. The National Hurricane Center (NHC) Tropical Cyclone Report for Hurricane Charley described a central pressure drop from 965 mb (1403 UTC) to 941 mb (1957 UTC) and an intensity increase from 90 kt (105 mph) to 125 kt (145 mph) within six hours. The official forecast track closely followed a 'consensus' model track forecast, and the forecast track error remained small. However, statistical and dynamical forecast aids did not capture the period of rapid intensification. While the official NHC forecast called for a strengthening storm (CAT 3 possible at landfall), the rate at which Charley deepened and the intensity it reached were not anticipated. In fact, only one intensity technique (Coupled Hurricane Intensity Prediction System—CHIPS) indicated a 120-plus knot storm. Whereas other available intensity guidance predicted a CAT 3 storm (or less) at landfall, the storm made landfall as a strong CAT 4.

2. Typhoon Nari (Okinawa, Japan 2001)

In the western North Pacific, Typhoon Nari (2001) proved a difficult track and intensity forecast. The JTWC Annual Tropical Cyclone Report 2001 (ATCR 2001) confirmed intensification (from 50 kt to 70 kt maximum sustained surface wind speed) over a six-hour period just before the storm made a second landfall

on Okinawa. Kadena Air Base near Naha, Okinawa, recorded a 66 kt sustained wind speed with a 99 kt gust. The TC track was even more problematic, since it had a series of cyclonic loops, with each loop bringing the storm center close to or across the island.

As the Nari example reveals, the potential for rapid intensification is certainly possible throughout the tropical western North Pacific, where forecasters at the JTWC are responsible for issuing tropical cyclone warnings for large portions of the Pacific (and Indian) Ocean basin(s). According to the JTWC Mission webpage, JTWC warnings are provided to all branches of the U. S. Department of Defense, and other U. S. government departments and their agencies such as the State Department (U.S Embassies and Consulates), and the Department of Commerce (U. S. National Weather Service). Moreover, forecasters at the JTWC must give tropical cyclone warnings for an array of U.S. military assets scattered throughout a tremendous area of responsibility, from the dateline westward to the African coast.

C. CHAPTER OVERVIEW

Whereas this introduction briefly describes TC forecasting challenges and stresses the importance of accurate TC prediction, the objective of this thesis is to evaluate TC intensity guidance accuracy using a conceptual intensity model for different stages of a TC life cycle. First, Chapter II provides some background on TC intensification and subsequent decay. Next, Chapter III introduces the idealized framework for TC intensity change analysis, summarizes storm information during the western North Pacific 2003 and 2004 tropical seasons, and introduces the TC intensity guidance techniques that are the focus of this thesis. The results from conventional intensity guidance evaluations as well as additional results from subjective model analyses are described in Chapter IV. Finally, Chapter V summarizes the key findings of this study and offers recommendations for future work. Moreover, this section describes the potential strengths and weaknesses of the intensity change techniques by giving

operational forecasters additional knowledge of where/when available intensity guidance may depart from reality.

II. TROPICAL CYCLONE INTENSIFICATION AND DECAY

A. TROPICAL CYCLONE POTENTIAL INTENSITY

Emanuel (1986, 1995a, 1995b, and 2004) and Holland (1997) examined the theoretical upper bound on TC intensity that might be achieved in conditions of minimal interaction between storms and their environment. Because favorable conditions for intensification must exist for some time after formation until the maximum intensity can be achieved, most of the time the storm is well below its maximum intensity. Nevertheless, the supposition is that an upper bound on TC intensity (minimum sea-level pressure (SLP) or maximum surface winds) can be calculated using one of these two methods.

Emanuel's technique uses the energy cycle of the storm to estimate maximum surface winds. Here, air parcels drawn toward the storm center acquire heat (high equivalent potential temperatures) from the ocean surface. As water vapor condenses during ascent, latent heat is converted to sensible heat, which results in observable temperature increases within the system. This 'deposition' of heat at high altitudes contributes to a decrease in surface pressure and a corresponding increase of surface wind speed (via gradient wind relation). An example of Emanuel's theory applied to real-time analyses of tropical basin maximum potential intensity (MPI, or simply PI) is given in Figure 2.1.

Holland (1997) expands upon research by Miller (1958). He proposes that the maximum temperature attainable in an eyewall may be estimated given the thermodynamic properties of the low-level storm inflow. In turn, this produces an estimate of the maximum temperature attainable in the eye owing to compressional warming of sinking air, and thereby leads to SLP and surface wind calculations. One can see that both of the Emanuel and Holland intensity limit calculations ultimately depend on sea-surface temperature (SST) and the vertical temperature structure of the atmosphere—variables that are calculable via regular data sets. Of greater note, both PI approaches yield similar results. Nevertheless, research suggests PI calculations work well as a *limit* to intensity,

but do not work well as a predictor of individual storm intensity (Emanuel webpage reference, cited 2005).

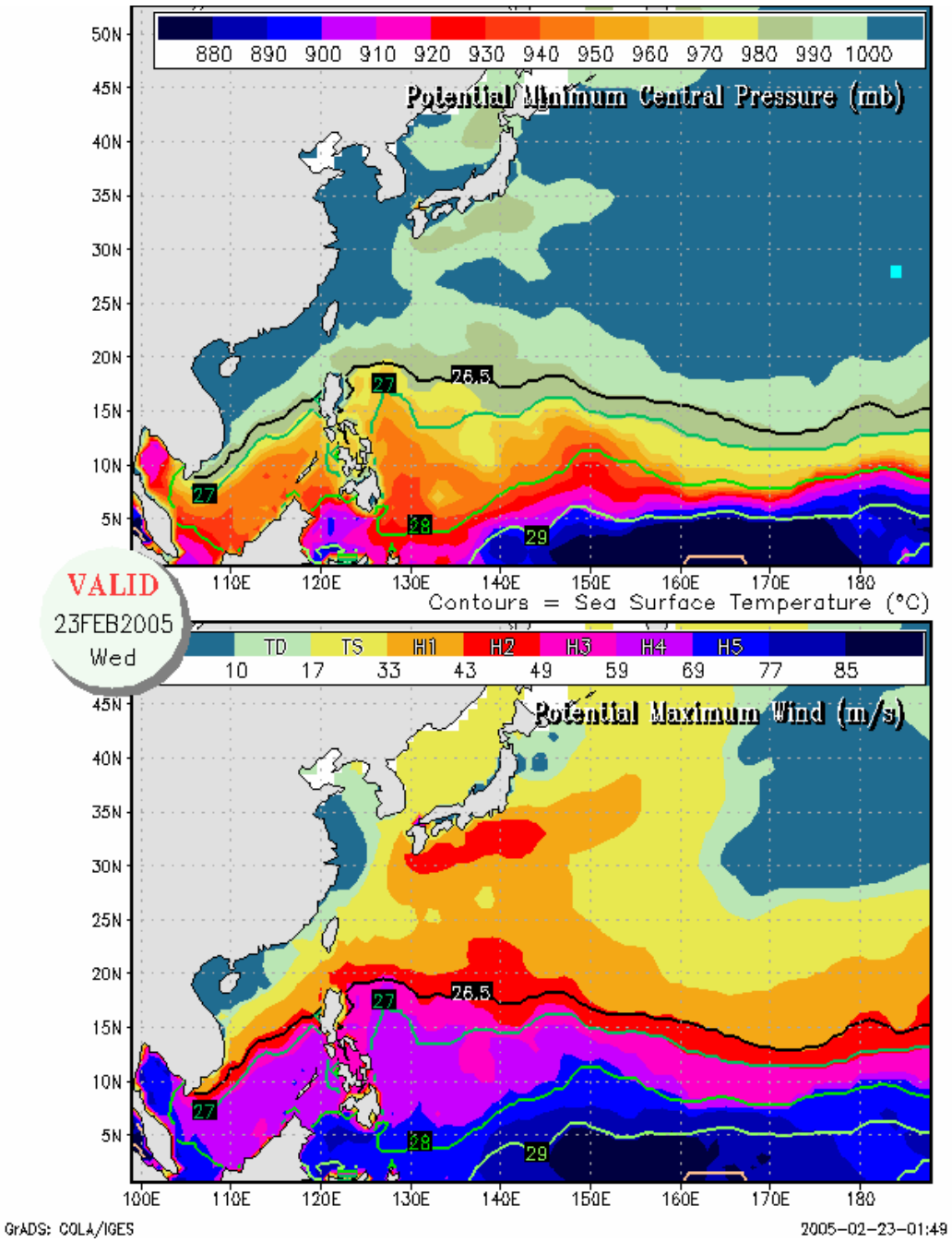

Figure 2.1 Potential intensity diagrams indicating minimum sea-level pressure (mb) and maximum attainable wind speed (m/s). The images were created by the Center for Ocean-Land-Atmosphere Studies (COLA) using Emanuel's PI theory. The maps are based on data from the 00UTC global operational analysis from the National Centers for Environmental Prediction (NCEP) for the date shown on the plot.

B. INTERNAL PROCESSES

Internal storm processes (e.g., concentric eyewall cycles) also contribute to intensity fluctuations. In some tropical systems, the result of these complex dynamic/thermodynamic interactions is the formation of multiple eyewalls. Willoughby et al. (1980) describe concentric convective rings as being most common in intense typhoons (winds greater than 130 kt), and their initial appearance usually marks the end of a strong intensification period. In the concentric eyewall process, a second (there are even cases of a third) eyewall forms well beyond the original radius of maximum winds. It is important to note that concentric eyewalls are not observed in storms below typhoon strength.

One question to be addressed in this thesis is whether available guidance can forecast a distinct eyewall replacement cycle. The concentric eyewall cycle affects the intensity in a series of steps: (i) a secondary eyewall begins to develop and an increase in wind speed is noted on the inward side of this newly-developed (secondary) wind maximum; (ii) as the outer ring constricts around the existing eye, the convective rings begin to interact; (iii) the innermost eyewall weakens (intensity decreases) while the secondary (outer) eyewall contracts (rather symmetrically) toward the center; and (iv) the intensity increases again as the secondary eyewall radius shrinks. Once the inner eye is completely gone, some reintensification is possible as the new eyewall contracts, often leading to a 'secondary peak' in storm intensity. While detection of multiple eyewall features has increased with improvements in satellite remote sensing (e.g., microwave imagery), forecasting the timing of intensity fluctuations remains tenuous.

C. MIDLATITUDE INTERACTIONS

Tropical cyclones also undergo dramatic intensity fluctuations due to complicated dynamic and thermodynamic interactions as the storm encroaches upon the midlatitudes (or when a midlatitude trough penetrates deep into the subtropics). Klein et al. (2000) describe the extratropical transition (ET) of western North Pacific storms. The authors also define two stages in the ET process: (i) 'transformation' in which the weakening TC evolves into an

asymmetric baroclinic storm; and (ii) 'reintensification' in which the transformed storm deepens as an extratropical cyclone. Such scenarios were observed in the 2003/2004 western North Pacific tropical seasons, as the JTWC best track data abruptly ends for poleward-moving storms undergoing ET (and also for storms making landfall).

The transformation stage begins as a poleward-moving TC interacts with a pre-existing baroclinic zone, which initiates an equatorward flow of colder, drier (midlatitude) air west of the storm center. This flow of drier air effectively suppresses deep convection (on the west side), and a dry slot forms in the southwestern quadrant. Conversely, poleward flow continues east of the storm center, advecting warm, moist air up isentropic surfaces within the pre-existing baroclinic zone. Continued advection on the east and west sides of the TC leads to a southwest to northeast orientation of the baroclinic zone. Descent in the western/southwestern portion of the storm compensates for the continued ascent east of the storm center. Furthermore, increased vertical wind shear associated with the baroclinic zone removes the upper-tropospheric warm core downstream, and the upper-tropospheric outflow becomes confluent with the polar jet stream. With continued poleward motion, the weakened storm center encounters lower SSTs and begins to resemble a midlatitude cyclone. In fact, Harr and Elsberry (2000) indicate warm frontogenesis north and east of the transformed storm center and weak cold frontogenesis south and west of the center. Finally, transformation is complete once the storm becomes embedded within the baroclinic zone, with the surface storm center lying in the relatively cold, descending air. It should also be noted that some storms never complete ET and continue to weaken while traversing lower SSTs (strong north-south SST gradient) and/or when encountering increased vertical wind shear.

Reintensification is a complex interaction between the environment and the extratropical storm, whereby the central pressure of the reinvigorated system deepens, and in some cases becomes stronger (lower central pressure) than the original tropical cyclone. Case studies by Harr et al. (2000) showed that these systems undergo the greatest deepening when they couple with a midlatitude

trough northwest of the circulation center (called a northwest pattern) via constructive fluxes of momentum and vorticity. Another less vigorous intensification pattern occurs when the trough northwest of the extratropical storm remains relatively weak, and the system instead is advected northeast toward a stronger, nearly-stationary (occluded) extratropical system. In this northeast pattern, the coupling between the midlatitude trough and the ex-TC is reduced, and the storm may or may not significantly intensify. It should be noted that once the JTWC classifies a storm as 'extratropical' the storm intensity is no longer verified—despite what could be a very potent, reintensifying storm with near-typhoon-force winds, active seas, and intense rainfall rates. The storm then becomes the responsibility of other regional/national forecast centers.

D. OTHER FACTORS

There are also many synoptic constraints on the different stages of TC intensity change. The original SAFA rationale was that the track and intensity forecasting should be integrated (Carr and Elsberry 1994). That is, the track affects intensity changes (e.g., over land, over SST gradients, and in different vertical wind shear regimes), and the TC intensity ultimately affects the depth of the steering layer. In general, outer wind structure was assumed to not change much on the time scale of 72-h track forecasts, so that much of the SAFA focus was on the storm structure at the initial time.

The effects of movement over land, different SST gradients, and vertical wind shear regimes are dependent on forecast track accuracy, which is especially true as a TC moves over land or just skirts a landmass. In this example, an erroneous track forecast over land would yield a large intensity forecast error, and vice versa if the TC does move over land and the intensity forecast is based on an official track forecast that the TC will remain over the ocean. Additionally, movement over lower SSTs will have a different effect on a weak TC than it does on an intense, mature TC. This SST effect is also related to the effect of vertical wind shear on the TC intensity. That is, the TC may be able to resist a larger amount of vertical wind shear if it remains over high SSTs,

but the combination of moving over lower SSTs and being exposed to increasing vertical shear (e.g., following recurvature) may lead to rapid decreases in wind speed. Thus, the same amount of vertical wind shear when the TC is in the developing stage may have a different intensity change than when it is in the decaying stage.

III. METHODOLOGY

A. FRAMEWORK FOR INTENSITY CHANGE ANALYSIS

Consider the conceptual model of the TC life cycle in Figure 3.1 as a framework for intensity change analysis. Tropical cyclones in the western North Pacific can form in different environmental conditions depending on the pre-TC seedling (e.g., monsoon depression or waves at the east end of the monsoon trough). These pre-TC seedlings may exist for days without changing intensity until moving into a favorable environment and formation (e.g., defined at 25 kt or 35 kt in Figure 3.1) occurs. The next critical stage in storm evolution following formation is intensification (Phase II in Fig. 3.1), where forecasters must decide the magnitude of intensity change and the duration of intensification. During this Phase II, track direction and speed influence intensification; e.g., a slow track has a longer time to intensify over warm water. The intensification in Phase II may also depend on the synoptic pattern/region and be quite different for a TC in the Standard/Tropical Easterlies from a TC in Poleward/Poleward Flow or Poleward/Equatorward Flow (see Carr and Elsberry 1994). The next critical forecast decision is determining the maximum intensity that will be achieved, which again is a function of the track direction and speed since the TC needs a certain length of time to approach its MPI for a given SST and atmospheric profile. In the simple conceptual model in Figure 3.1, a final critical forecast decision is determining the onset and magnitude of TC decay (Phase III). As indicated in Chapter II D, this Phase III is critically dependent on the track over land or over decreasing SST and increasing vertical wind shear.

The discussion of internal processes in Chapter II B, such as the eyewall replacement cycles, makes the conceptual model too simple. Thus, a more general conceptual model of the life cycle is introduced in Figure 3.2, in which Phase IIa includes the possibility of one or more decay/intensification cycles. Therefore, the forecaster must consider whether the storm will have intensity fluctuations after reaching peak intensity.

Guidance for each of these critical forecast decisions may be examined using the framework of the idealized intensity profiles in Figures 3.1 and 3.2. That is, do the intensity forecast techniques provide accurate guidance during these stages in a typical life cycle: (i) formation, defined here as a pre-TC seedling through tropical depression (TD) status; (ii) intensification, described as early intensification through tropical storm (TS) status and often followed by a period of more rapid intensification leading up to peak intensity; (iii) intensity fluctuations following peak intensity, often due to reintensification following an eyewall replacement cycle; and (iv) the primary period of decay, defined either as a steady decay following a single peak intensity (Figure 3.1), or the final storm decay following any decay/reintensification periods (Figure 3.2)? This thesis primarily examines the intensity changes following TC formation, since this is when the magnitude of intensity changes is greatest. Additionally, many of the intensity guidance techniques do not provide intensity guidance during the formation period, as the storm is often too weak for the models to capture this phase.

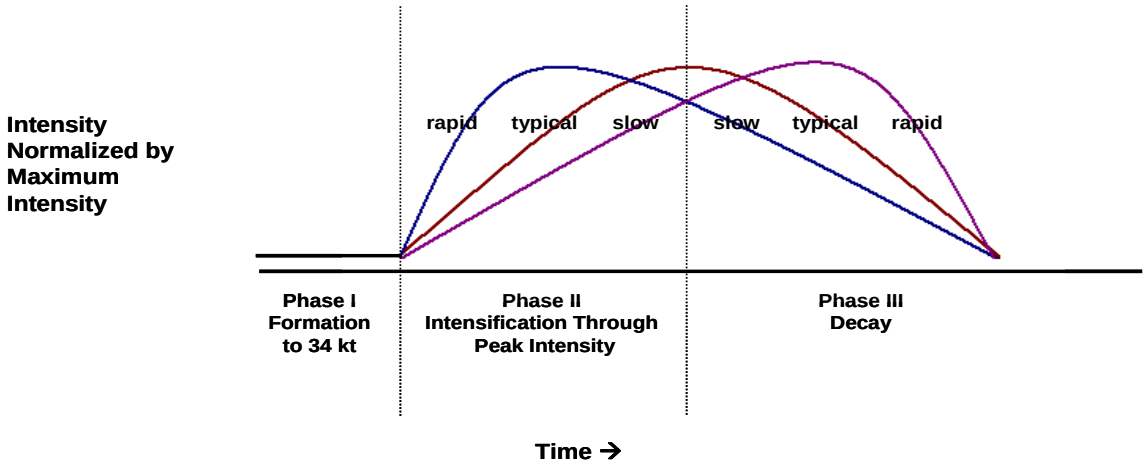

Figure 3.1 Idealized intensity traces showing intensification and decay profiles following the formation phase (Phase I). Phase II intensification may be described as slow (modest increase in intensity for a given time period), typical (an average rate of intensification for a given ocean basin and time period), or rapid (an above-average rate of intensification for a given time period). Phase III may be described in a similar manner: slow, typical, and rapid rates of decay.

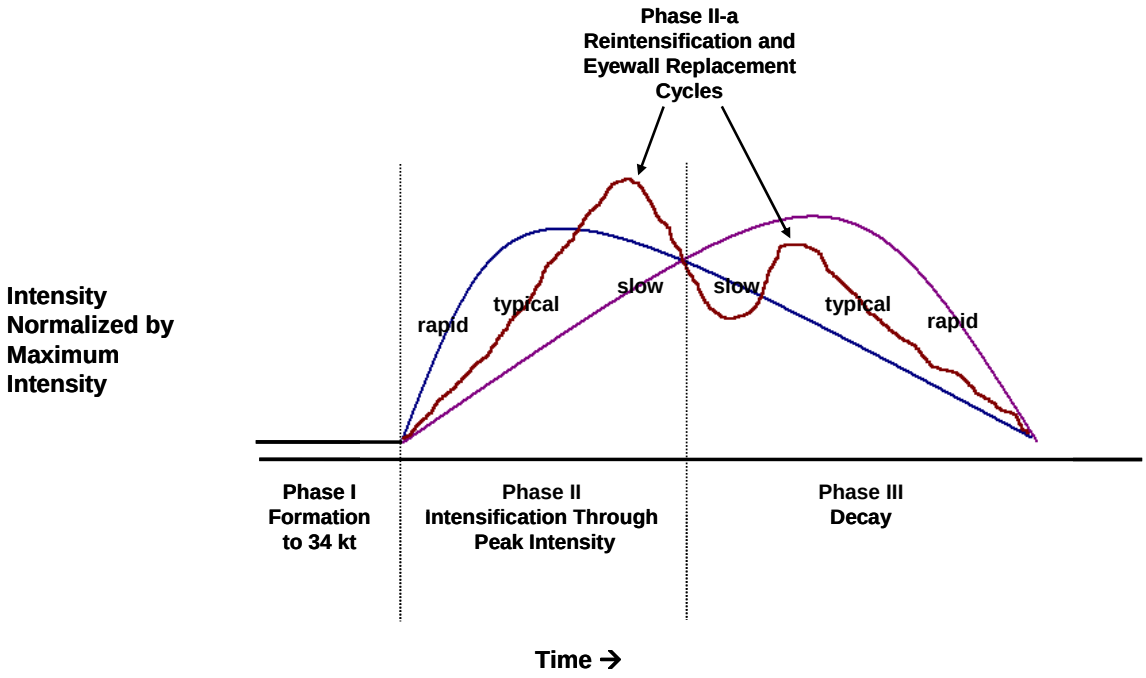

Figure 3.2 Same as Fig. 3.1, except with the addition of Phase IIa to represent reintensification following an initial period of decay, especially during eyewall replacement cycles observed in intense TCs.

B. DATA SOURCES

The JTWC aids files, which include best track information and the various intensity forecasts, are the primary data sources for the 2003 and 2004 intensity forecast evaluations. These files are part of the Automated Tropical Cyclone Forecasting System (ATCF) developed by the Naval Research Laboratory, Monterey, CA. Additional storm-specific information was gleaned from the JTWC 2003/2004 Annual Tropical Cyclone Report (ATCR) and 2003 and 2004 North Pacific TC Summaries (courtesy of Gary Padgett at tropical-storms-admin@tstorms.org). Ancillary storm details were derived from the 'Digital Typhoon' web-based resource of the Japan Meteorological Agency (JMA) [Available at http://agora.ex.nii.ac.jp/digital-typhoon/ (Current as of 15 Feb 05)].

C. FRAMEWORK APPLIED TO THE 2003 AND 2004 SEASONS

A total of 59 storms (see storm summaries for the 2003 and 2004 seasons in Tables 3.1 and 3.2) were used to apply the framework for intensity analysis. First, 'storm duration' is defined as the number of days a storm existed from initial

formation (Phase I) through final decay (Phase III); e.g., between the first best track intensity report and the last best track intensity report in the JTWC warnings. The durations of the storms were grouped into three classes: (i) short-lived storm (lasting six days or less); (ii) an averaged-lived storm (lasting greater than six days to 12 days); (iii) or a long-lived storm (lasting greater than 12 days). These categories were defined using natural breaks in the distribution of storm duration (Fig. 3.3).

TC	Name	Month	Warning Duration (days)	Est. Max. SFC Wind (kt)	Formation Belt (LAT)	Formation Belt (LON)	Track
TS 01W	YanYan	Jan	10.25	50	5N to 9N	East of 145E	Recurve
STY 02W	Kujira	Apr	17.25	135	0 to 4N	East of 145E	Recurve
TD 03W		May	3.5	30	5N to 9N	120E to 145E	Erratic N
TY 04W	Chan-hom	May	9.25	115	5N to 9N	East of 145E	Recurve
TS 05W	Linfa	May	6.5	60	15N to 19N	West of 120E	NE
TS 06W	Nangka	May-Jun	4	50	15N to 19N	West of 120E	NE
TY 07W	Soudelor	Jun	13	115	5N to 9N	120E to 145E	Recurve
TY 08W	Koni	Jul	8.25	65	10N to 14N	120E to 145E	WNW-NW
STY 09W	Imbudo	Jul	9.5	130	5N to 9N	East of 145E	NW
TY 10W	Morakot	Aug	5.25	65	15N to 19N	120E to 145E	NW
TY 11W	Etau	Aug	9.5	110	10N to 14N	120E to 145E	Recurve
TY 12W	Krovanh	Aug	12.25	90	5N to 9N	East of 145E	NW-W
TS 13W	Vamco	Aug	1.5	35	15N to 19N	120E to 145E	N-NW
TY 14W	Dujuan	Aug-Sep	8	125	15N to 19N	120E to 145E	SW-WNW
STY 15W	Maemi	Sep	10	150	10N to 14N	120E to 145E	Recurve
TY 16W	Choi-Wan	Sep	5.5	95	10N to 14N	120E to 145E	Recurve
TY 17W	Koppu	Sep	8	90	15N to 19N	120E to 145E	Recurve
TD 18W		Oct	4.5	25	15N to 19N	West of 120E	Erratic NW
TD 19W		Oct	3.5	30	25N to 29N	120E to 145E	NE
TY 20W	Ketsana	Oct	9.5	125	15N to 19N	120E to 145E	NE
STY 21W	Parma	Oct	13.25	130	15N to 19N	120E to 145E	C.W. Loop
TD 22W		Oct	2.75	25	10N to 14N	West of 120E	E
TS 23W		Oct	6.75	35	10N to 14N	West of 120E	WNW
TY 24W	Melor	Oct-Nov	6.25	70	10N to 14N	West of 120E	Recurve
TY 25W	Nepartak	Nov	8.25	75	10N to 14N	120E to 145E	W-NW
STY 26W	Lupit	Nov-Dec	16.5	145	5N to 9N	East of 145E	Recurve
TS 27W		Dec	6	35	10N to 14N	120E to 145E	W-SW

Table 3.1 Storm information adapted from the JTWC Annual Tropical Cyclone Report (ATCR) 2003. Storm names were designated by the Regional Specialized Meteorological Center (RSMC), Tokyo, or by the Central Pacific Hurricane Center (CPHC), Honolulu. The JTWC provided warnings for 27 tropical cyclones during the 2003 season: four tropical depressions; six tropical storms; and 17 typhoons (five super typhoons with estimated sustained surface wind speed equal to or exceeding 130 kt).

TC	Name	Month	Warning Duration (days)	Est. Max. SFC Wind (kt)	Formation Belt (LAT)	Formation Belt (LON)	Track
TS 01W		Feb	10.5	45	10N to 14N	120E to 145E	C.C.W. Loop
TS 02W		Mar	9.75	45	5N to 9N	120E to 145E	NW
STY 03W	Sudal	Apr	15	130	5N to 9N	East of 145E	Recurve
STY 04W	Nida	May	8.25	140	5N to 9N	120E to 145E	Recurve
TS 05W		May	7.75	35	5N to 9N	West of 120E	Short Erratic
TY 06W	Omais	May	8.25	65	5N to 9N	120E to 145E	WNW - NE
TY 07W	Conson	Jun	13	85	15N to 19N	West of 120E	NE
TY 08W	Chanthu	Jun	8.25	75	10N to 14N	West of 120E	WNW
STY 09W	Dianmu	Jun	10.25	150	5N to 9N	120E to 145E	Recurve
TY 10W	Mindulle	Jun-Jul	14.5	125	15N to 19N	120E to 145E	Recurve
TY 11W	Tingting	Jun-Jul	11.5	80	10N to 14N	East of 145E	Recurve
TS 12W	Kompasu	Jul	6	45	20N to 24N	120E to 145E	W
TY 13W	Namtheun	Jul-Aug	8.5	115	20N to 24N	East of 145E	NW - WNW
TY 14W	Meranti	Aug	5.75	90	20N to 24N	East of 145E	NE - N
TD 15W	Malou	Aug	5	30	30N to 34N	120E to 145E	NW - N
TY 16W	Rananim	Aug	3.5	90	15N to 19N	120E to 145E	N - NW
TS 17W	Malakas	Aug	2.25	35	25N to 29N	East of 145E	NE
TY 18W	Megi	Aug	7.25	65	15N to 19N	120E to 145E	Recurve
STY 19W	Chaba	Aug	14.25	155	10N to 14N	East of 145E	Recurve
TY 20W	Aere	Aug-Sep	13.5	65	10N to 14N	120E to 145E	NW - WSW
TS 21W		Aug	4.25	35	15N to 19N	East of 145E	WNW
STY 22W	Songda	Aug-Sep	12.75	130	10N to 14N	East of 145E	Recurve
TS 23W	Sarika	Sep	4.5	60	15N to 19N	East of 145E	WNW - N
TD 24W	Haima	Sep	3.5	30	20N to 24N	120E to 145E	NNE - NW
TY 25W	Meari	Sep	12	120	10N to 14N	120E to 145E	Recurve
STY 26W	Ma-on	Oct	11.75	140	15N to 19N	120E to 145E	Recurve
TY 27W	Tokage	Oct	11	120	10N to 14N	East of 145E	Recurve
TY 28W	Nock-ten	Oct	10	110	10N to 14N	East of 145E	WSW - NW
TY 29W	Muifa	Nov	13.75	85	10N to 14N	120E to 145E	NW - SW
TY 30W	Nanmadol	Nov-Dec	5.75	125	5N to 9N	East of 145E	NW
TS 31W	Talas	Dec	11	40	5N to 9N	East of 145E	W - NNW
TS 32W	Noru	Dec	5	55	10N to 14N	East of 145E	Recurve

Table 3.2 Same as Table 3.1, except for the 2004 western North Pacific tropical season. A total of 32 tropical cyclones occurred during 2004: two tropical depressions; nine tropical storms; and 21 typhoons (six super typhoons).

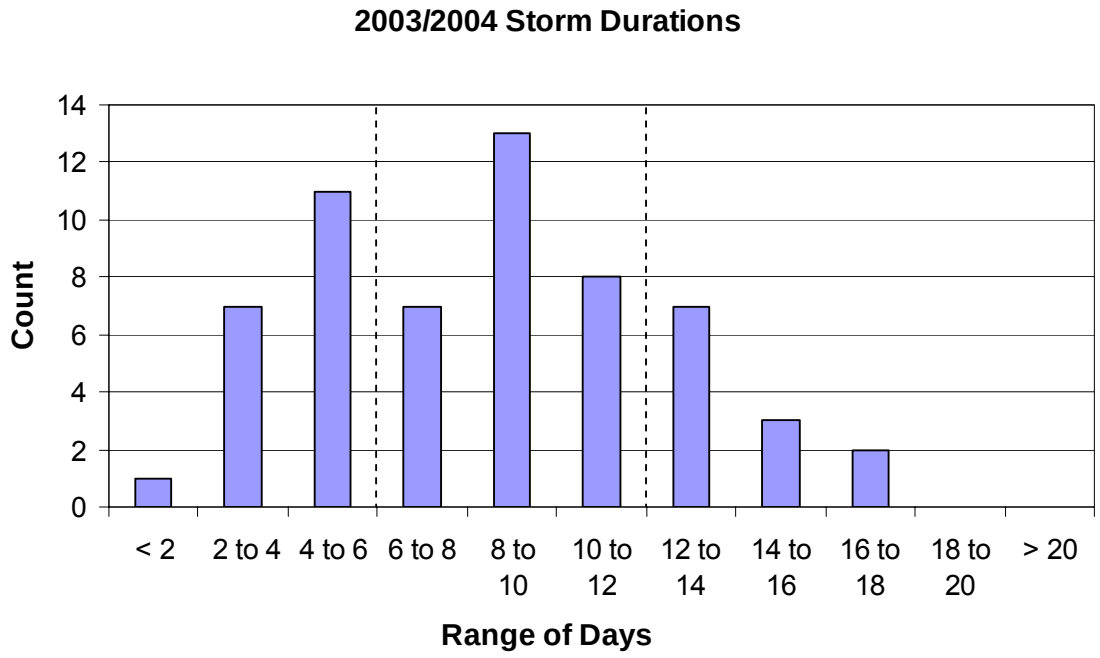

Figure 3.3 Combined 2003/2004 storm durations (life spans).

A summary of storm durations for the 2003 and 2004 seasons breaks down as follows: 19 of 59 storms were short lived; 28 of 59 storms were averaged lived; and 12 of 59 total were long lived. It is likely that the short-lived storms will have a life cycle similar to Figure 3.1, and average-lived (or especially long-lived) storms are more likely to have a life cycle similar to Figure 3.2 with a decay and re-intensification cycle.

- The 19 short-lived TCs included six tropical depressions, eight tropical storms, and five typhoons ;
- The 28 averaged-lived TCs included zero tropical depressions, seven tropical storms, and 21 typhoons (five super typhoons);
- The 12 long-lived TCs were all typhoons (with 6 of the 12 typhoons reaching super typhoon status).

While storm duration is simply the total number of Phase I through Phase III days, 'phase duration' is defined as the total number of days a particular storm spends in each phase of the intensity framework (Phases I, II, or III in Figures 3.1 and 3.2). Therefore, Phase I duration is defined as the total time the storm spends in the formation phase until surpassing 34 kt (the tropical storm

threshold). Phase II duration (and Phase II-a) is defined as the total time the storm spends intensifying until reaching peak intensity (or a secondary peak in intensity). Finally, Phase III duration is defined as the total time the storm spends in the primary decay phase (not including Phase II-a intensity fluctuations). The combined 2003 and 2004 phase duration averages follow:

- Short-lived storms: Phase I (2.5 days); Phase II (1.5 days); Phase III (0.75 days).
- Average-lived storms: Phase I (3.0 days); Phase II (3.5 days); Phase III (3.0 days).
- Long-lived storms: Phase I (4.25 days); Phase II (7.0 days); Phase III (3.25 days).

Note that a pre-TC seedling may exist in Phase I for 2.5-4.25 days on average before achieving TS status. During this Phase I, the intensity guidance should be predicting near-constant intensity or modest intensification. Once into the intensification (Phase II) the short-lived storms have a shorter period to reach maximum intensity than do longer-lived storms. Likewise, the decay period (Phase III) is much shorter for the short-lived storms than for average-lived and long-lived storms. Also note that the longer-lived storms spend more time in each of the three phases than the shorter-lived storms.

If one were to combine these phase duration averages with intensity information, an expected rate of intensification or decay could be calculated. For example, maximum storm intensity averaged for the 59 storms during the 2003 and 2004 seasons was about 85 kt (rounded to the nearest five knots). This intensity average may also be examined as intensity averages for short-lived storms (50 kt), average-lived storms (90 kt), and long-lived storms (115 kt). Therefore, if one were to apply these intensity averages to the Phase II durations of short-, average-, and long-lived storms, one could reasonably expect the following intensification rates:

- Short-lived storms: 10 kt/day;
- Average-lived storms: 15 kt/day;
- Long-lived storms: 20 kt/day.

The same procedure may be applied for decay, with the exception of choosing a 45 kt decay threshold (since the best track data may not continue decay past 45 kt for many storms due to ET or landfall). Therefore, if one were to apply these intensity averages to the Phase III durations of short-, average-, and long-lived storms, one could reasonably expect the following decay rates:

- Short-lived storms: 5-10 kt/day;
- Average-lived storms: 15 kt/day;
- Long-lived storms: 20-25 kt/day.

While these simple evaluations may miss periods of slower or more rapid rates in a particular intensification/decay cycle, it provides a 'first-guess' for comparison with the available intensity change guidance. These calculated (average) intensification rates fall within the observed intensification rates of many of the intensity guidance techniques. The characterizations of the intensification/decay rates (i.e., rapid, typical, and slow) will be established from comparisons between the different model intensity change forecasts and the actual rates of intensification/decay.

Thus, the intensity framework in Figure 3.1 and 3.2 can be useful in distinguishing the intensity characteristics of different storms. Suppose the forecaster can determine from the formation location and expected track that the pre-TC seedling will at most be a short-lived storm. Then the expected intensification rate during Phase II, maximum intensity to be achieved, and decay rate during Phase III are going to differ from the average-lived storm. Likewise, if the formation is deep in the tropics and a long path in favorable conditions is expected, the storm is more likely to approach the PI, more likely to have at least one decay/reintensification cycle (Phase IIa), and to have a longer primary decay period (Phase III) than the short-lived storm.

D. SELECTED INTENSITY GUIDANCE TECHNIQUES

Since this thesis focuses on evaluations of intensity guidance accuracy, some basic knowledge of the various intensity guidance techniques is necessary. Moreover, the performance of these techniques will be evaluated during each

phase of the intensity framework described above. The current pool of readily available, operational TC intensity guidance techniques for the western North Pacific exceeds 10 techniques (Appendix A). Although a comprehensive comparison of all operational intensity guidance is beyond the scope of this thesis, three primary intensity guidance techniques are considered: *statistical, statistical-dynamical, and dynamic*. Each of the three types of techniques offers different approaches to TC intensity forecasting: each technique requires a different design; each technique utilizes a different set of initial conditions; each technique utilizes a distinct set of predictive equations; and thus each technique is expected to have unique model characteristics. These differences often lead to diverging intensity predictions and sometimes large deviations from the observed storm intensity.

This thesis evaluates the performance from six of the techniques for the western North Pacific tropical cyclone seasons of 2003 and 2004: Statistical 5-Day Typhoon Intensity Forecast—5-day Model (ST5D); Statistical Typhoon Intensity Prediction Scheme (STIP); Geophysical Fluid Dynamics Model–Navy (interpolated) (GFNI); Air Force Weather Agency Mesoscale Meteorological Model-5 (interpolated) (AFWI); Coupled Hurricane–Ocean Intensity Prediction System (CHIP); and Japan Typhoon Model (interpolated) (JTYI). Official intensity forecasts (JTWC) are also evaluated for comparison.

First, the ST5D represents a statistical intensity forecasting approach that requires only climatological data and the present intensity, recent intensity change, and a few storm variables to predict future TC intensity (see Appendix B for details). The STIP combines elements of the ST5D statistical approach with dynamically-produced model fields, and thus may be characterized as a statistical-dynamical model for TC intensity forecasting (see Appendix C for details). The CHIP provides an example of a relatively simple dynamic TC intensity prediction model because it moves an axisymmetric TC model over a one-dimensional ocean mixed-layer model (see Appendix D for details). Thus, the CHIP model combines the current hurricane/ocean storm environment and predicts future intensity changes along the official forecast track. Three

sophisticated, full physics models that provide track *and* intensity guidance are also considered: GFNI (see Appendix E for details); AFWI (see Appendix F for details); and JTYI (see Appendix G for details).

Note that the interpolated versions of the three full physics models are used. These interpolated techniques were chosen due to operational considerations, as the interpolated output is available prior to the end of the model computational cycle. Therefore, if forecasters wish to maximize the number of available intensity guidance techniques, the interpolated output must be used. In contrast, processing time for the ST5D, STIP, and CHIP techniques is rather short, and intensity change guidance is available to forecasters in a fraction of the time required to generate output from the full-physics techniques.

IV. RESULTS

A. BASIC STATISTICS

The conventional analysis of intensity forecasts is to calculate a simple average of the (forecast – observed) intensity bias (positive is an over forecast) and an average of the absolute value of the (forecast – observed) intensity difference. The latter metric is then a typical magnitude of error, with no special penalty for larger errors as is implicit in a root mean square error in which the errors are first squared before averaging. In the following discussion, these will be referred to as average intensity errors. In each of the inter-comparisons of the intensity techniques, a homogeneous sample is used (i.e., exactly the same forecasts are being compared, which is necessary because some techniques are not available every time and some forecasts are easier than others are. The lone exception to the homogeneous evaluation is the JTYI model. This model required a separate basic statistical analysis, and the results were later combined with the homogeneous data set for comparison.

1. Average Bias and Error Results for 2003

If the technique had as many over-forecasts as under-forecasts, the bias of the intensity forecasts should be near zero in a large sample. While this bias was near zero at time zero for all of the techniques except CHIP, the biases generally increased with each forecast interval (Fig 4.1). The tendency for CHIP is to have an increasingly large negative bias (under-forecasting intensity) during 2003, with a 72-h bias of –38 kt. The climatology and persistence technique ST5D had a negative bias of about –13 kt at 72 h, which suggests the 2003 cyclone sample did not have a similar intensity change distribution as for the developmental sample on which ST5D is based. Surprisingly, the statistical-dynamical technique STIP had an even larger negative bias (~ -25 kt) at 72 h. Whereas the STIP includes the ST5D predictors, which may account for some of the bias, one would expect the other predictors in STIP to have reduced the bias. The dynamical models had smaller biases, with the GFNI and AFWI models being positive (tendency to over-forecast intensity) and the JTYI being negative.

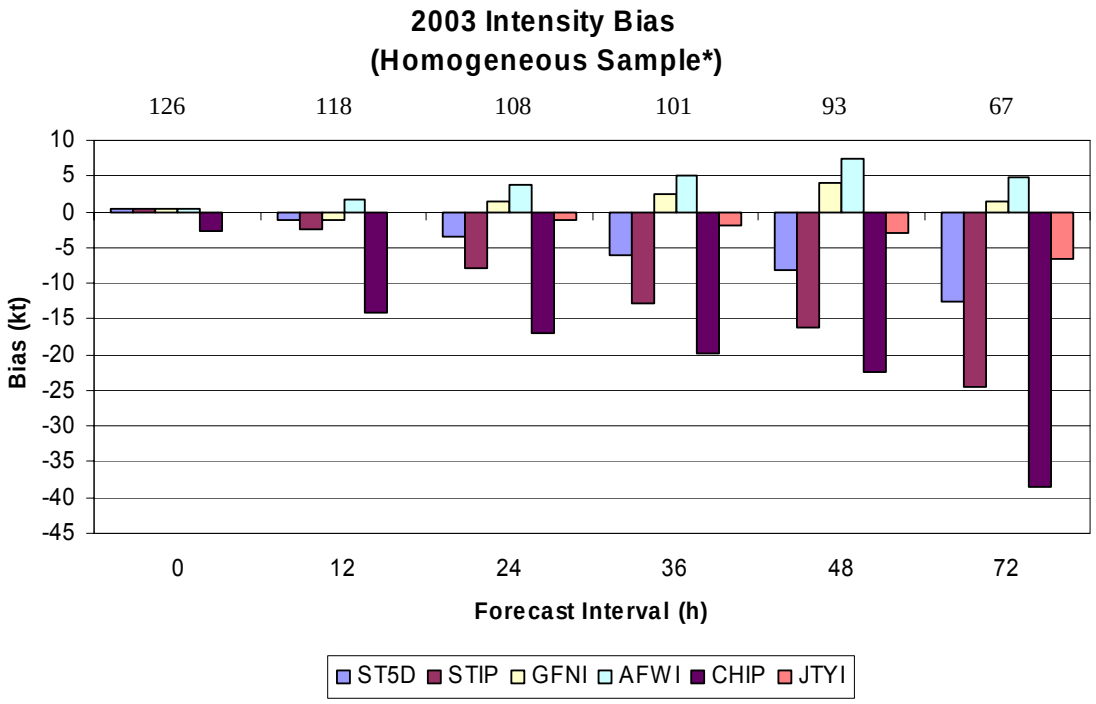

Figure 4.1 Bias of the intensity forecasts by six techniques (see box for colors) at analysis time (time = 0) and at 12, 24, 36, 48, and 72 h during the 2003 season. Numbers along the top of the graph represent the sample sizes of homogeneous cases. Note the * indicates that the JTYI is a non-homogenous member used for comparison.

Note that the average magnitude of intensity error (Fig. 4.2) during the 2003 season includes the bias error and the random error. The average error increased with each forecast interval for all techniques, with the largest errors being associated with the AFWI and the CHIP models. This large error for CHIP is not surprising given the large bias displayed in Fig. 4.1. Since the bias-error for the AFWI model was small, a large average intensity error means both large under-forecasts and large over-forecasts are included. The ST5D is considered to be a minimal skill metric since it requires no special meteorological knowledge. By this metric, none of the five other techniques had any skill relative to ST5D during the 2003 season. Thus, the JTWC had little skillful guidance on average from these techniques.

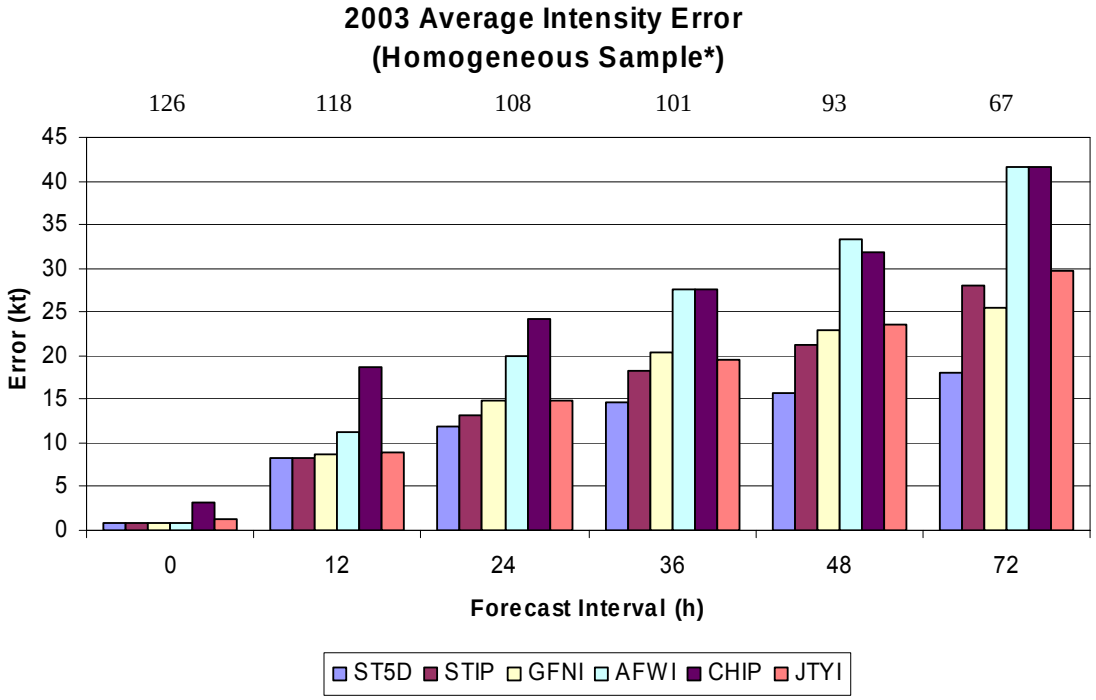

Figure 4.2 Average magnitude of the intensity forecast errors (kt) for six techniques at analysis time (time = 0) and at 12, 24, 36, 48, and 72 h during the 2003 season. Numbers along the top of the graph represent the sample sizes of homogeneous cases. Note the * indicates that the JTYI is a non-homogenous member used for comparison.

2. Average Bias and Error Results for 2004

The conventional bias analysis was repeated for the 2004 season (Fig. 4.3). Once again, the bias at zero hour was near zero with the exception of CHIP. The magnitude of the ST5D bias decreased slightly between 36 and 72 h and was of opposite sign from the 2003 season (from negative to slightly positive). The magnitude of CHIP biases was dramatically smaller at all forecast intervals during 2004 compared to 2003. This reduction was especially evident at the 72 h forecast interval, where the magnitude of the average bias decreased by 30 kt (from ~-38 kt in 2003 to ~-8 kt in 2004). Unlike 2003, when all of the techniques tended to have increasingly larger bias at longer forecast intervals, an increasing bias with longer forecast intervals was only evident in three of the models during 2004. First, the statistical-dynamical STIP had an increasingly negative bias with each forecast interval past 12 h. However, the magnitude of

the bias was smaller between 24 and 72 h when compared to the 2003 season. On the other hand, the GFNI and AFWI techniques had positive biases greater than the positive biases during 2003. Finally, the JTYI remained the technique with the smallest bias at nearly every forecast interval, which indicates a near-equal amount of over-forecasts and under-forecasts.

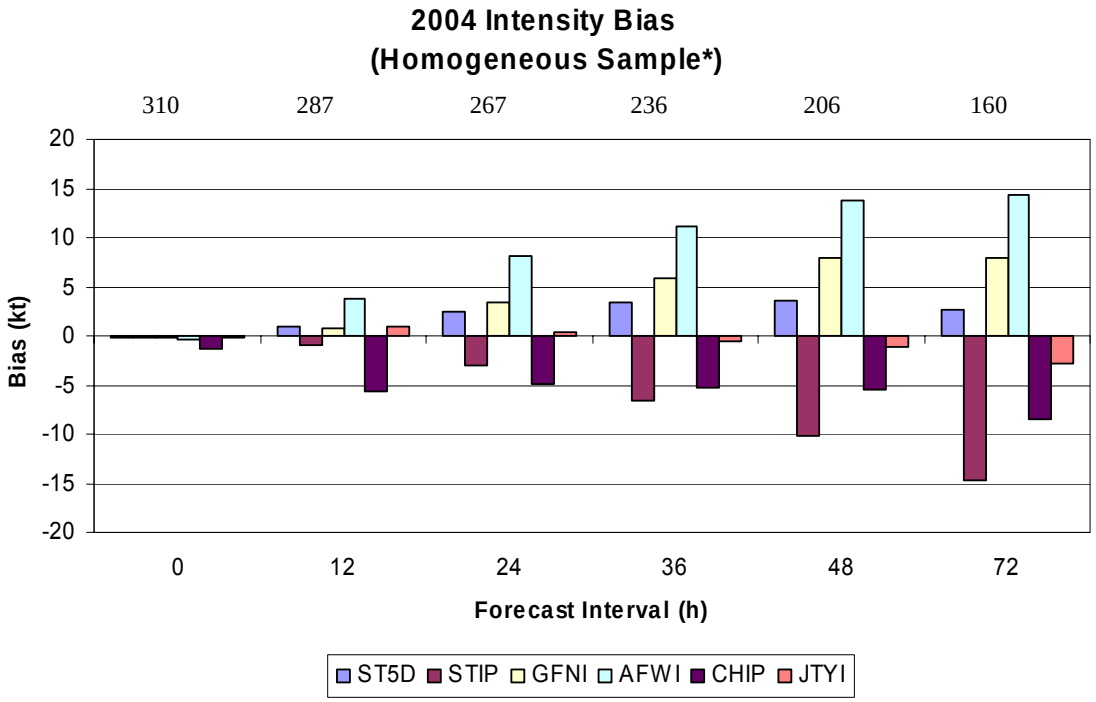

Figure 4.3 Bias of the intensity forecasts as in Fig. 4.1, except for the 2004 season. Numbers along the top of the graph represent the sample sizes of homogeneous cases. Note the * indicates that the JTYI is a non-homogenous member used for comparison.

As during 2003, the average intensity error during 2004 increased with each forecast interval for all techniques, with the largest errors beyond 24 h being associated with the AFWI and JTYI models (Fig. 4.4). The magnitude of intensity error associated with the STIP and GFNI models changed little from the previous season. While the CHIP model had very large errors during 2003, the magnitude of those errors dropped significantly during 2004. In fact, the CHIP model had the smallest intensity errors at 48 h and 72 h during 2004. Overall, the minimal skill ST5D had the smallest average intensity errors with the exception of CHIP at 48 and 72 h. That is, the only technique to add skill relative

to ST5D during the 2004 season was CHIP. Once again, the JTWC had little skillful guidance on average from these techniques.

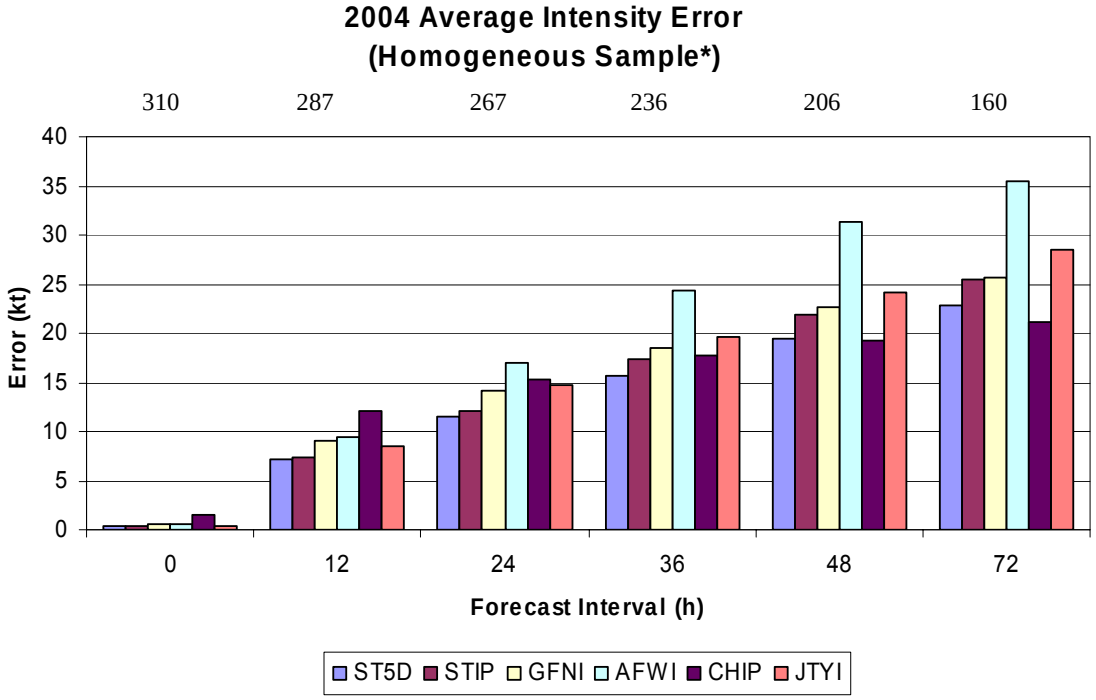

Figure 4.4 Average magnitude of the intensity forecast errors as in Fig. 4.2, except for the 2004 season. Numbers along the top of the graph represent the sample sizes of homogeneous cases. Note the * indicates that the JTYI is a non-homogenous member used for comparison.

B. PERFORMANCE DURING FORMATION PHASE

Recall from the intensity framework that the formation phase is a period of storm organization when anticipated intensification rates are rather slow (~10 kt/day or less). After all, the formation stage (as applied to this thesis) only lasts until the storm surpasses the 34-kt point, so there is not much opportunity for dramatic intensity increases. In fact, there were many occurrences when one (or more) of the intensity guidance techniques failed to produce intensity forecasts during Phase I. As a result, empirical evaluations of Phase I model performance proved difficult. Therefore, an objective evaluation tool was used to assess intensity guidance accuracy. Note that this same approach can be applied to all phases in the conceptual intensity framework.

A simple test of the usefulness of an intensity guidance technique is whether it at least indicates the intensity tendency over the forecast interval: increasing (> 10 kt); remaining nearly constant (+10 kt to -10 kt); or decaying (> -10 kt). These tendencies can be displayed in a contingency table as in Fig. 4.5, where the Good (G) cases are along the diagonal. With this procedure in mind, the intensity trend analyses can be calculated for all forecast intervals in the different phases of the intensity framework (Phase I to Phase III). That is, each analysis revealed how well a particular intensity prediction technique could forecast intensity change trends throughout the storm life cycle. Whereas a technique may have a Good (G) intensity trend performance (on average), it will still have cases of poor performance and the challenge is to determine when and why the guidance will be poor.

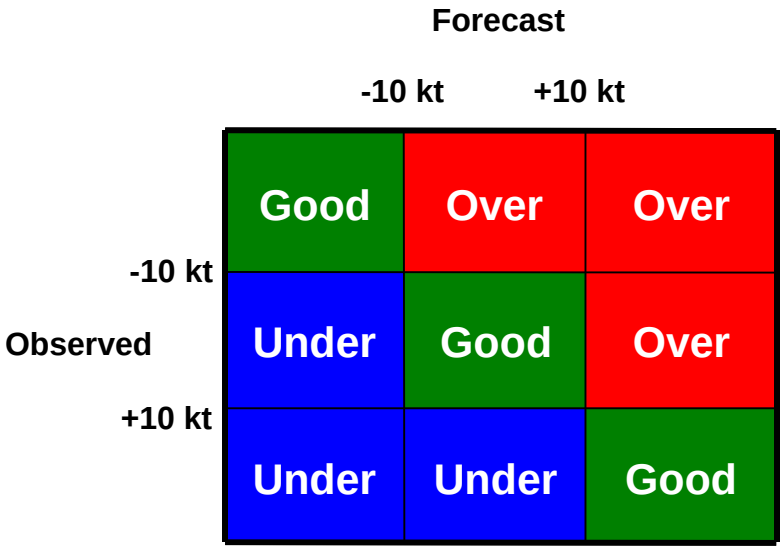

Figure 4.5 Sample contingency table with Good (G) intensity trend forecasts along the diagonal. A Good forecast was tallied if the observed intensity change fell within +/- 10 kt of the observed intensity verifying at each 12-h forecast interval; an Under (U) trend represents forecast decay greater than 10 kt, but the decay trend was not observed; an Over (O) forecast represents forecast intensification greater than 10 kt, but the intensification trend was not observed.

1. Good (G) Intensity Trends During Phase I

The first trend analysis (Fig. 4.6) examined model performance during the storm formation stage. All of the intensity techniques have a high percentage of

G intensity trends (80% or greater) for the 12-h forecast point. If the ST5D is used as a skill measure, none of the intensity techniques or JTWC has skill in this 12-h intensity trend test. However, the percentages of G intensity forecasts decreased dramatically by 24 h (~60% or less), and again none of the techniques or JTWC had skill relative to the ST5D. The STIP technique had the largest percentage of G intensity trend forecasts between 36 and 60 h, and was equal to the percentage for the JTYI at 72 h. Not many of the techniques besides ST5D and STIP provide 84-h and longer-range forecasts. For those forecasts that can be verified, the percentages of good forecasts are about 90%.

While the JTYI performance was relatively high (above 70%) at 60, 72, and 84 h, this technique generated only 40% G intensity trends at the 36-h point. In fact, intensity trend guidance from the JTYI was also poor at 24 and 48 h, as nearly one-half of all forecasts provided erroneous intensity change guidance during formation. The same assessment holds for the GFNI technique at 72 and 84 h, when over one-half of all forecasts generated poor guidance. If only 50% of the intensity technique predictions are not properly predicting the sense of the intensity trend to within 10 kt, then this is not useful guidance.

2. Under (U) Intensity Trends During Phase I

The percentage of U intensity trends generally increased between 12 and 36 h for all of the techniques (Fig. 4.7). This increase was especially evident in the JTYI and to a lesser extent in the GFNI, AFWI, and CHIP techniques. These rather significant U percentages indicate that these techniques may predict zero intensity changes or decreases even though the storm is intensifying during Phase I. Conversely, the relatively low percentages of under-forecasts by the ST5D and STIP techniques, except for the 24 h forecast interval, indicate that these techniques rarely produce under-forecasts of intensity change during formation (e.g., almost exclusively generate forecasts of intensification when the storm is still weak). In Figure 4.7, the measure of skill is that the techniques or JTWC should have smaller percentages of U forecasts than ST5D. Thus the skillful forecasts are: 12 h, GFNI and AFWI; 24 h through 48 h, STIP and JTWC; and 72 h through 120 h, none. As in the Phase I Good intensity trends above,

the sample sizes of forecasts verifying beyond 84 h drop considerably, which results in higher-than-expected U intensity trend percentages between 96 and 120 h. Given the small skill (too many U forecasts) by the various techniques, not much useful guidance for formation is available.

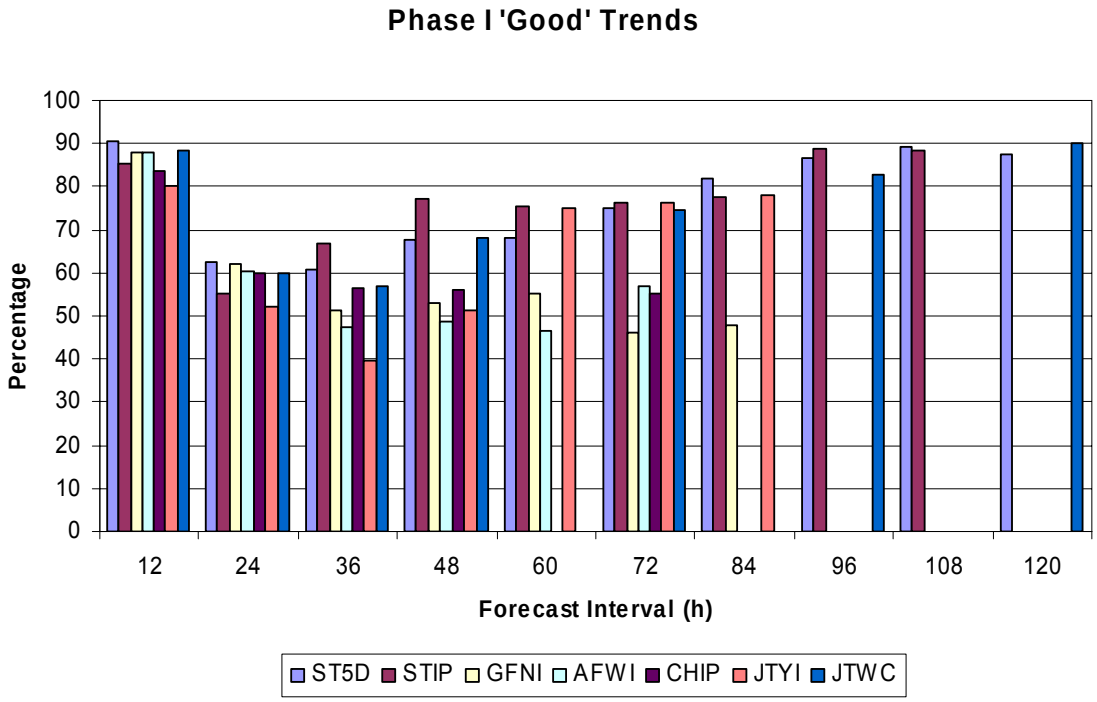

Figure 4.6 Percentage of Good intensity trends as defined in Figure 4.5 during storm formation (Phase I) for the combined 2003 and 2004 seasons. Sample sizes of the verified forecasts range from over 400 at 12 h to less than 30 by the 96-h point. A Good forecast during formation indicates that the magnitude of the forecast intensification rate is within +/- 10 kt of the actual intensification rate.

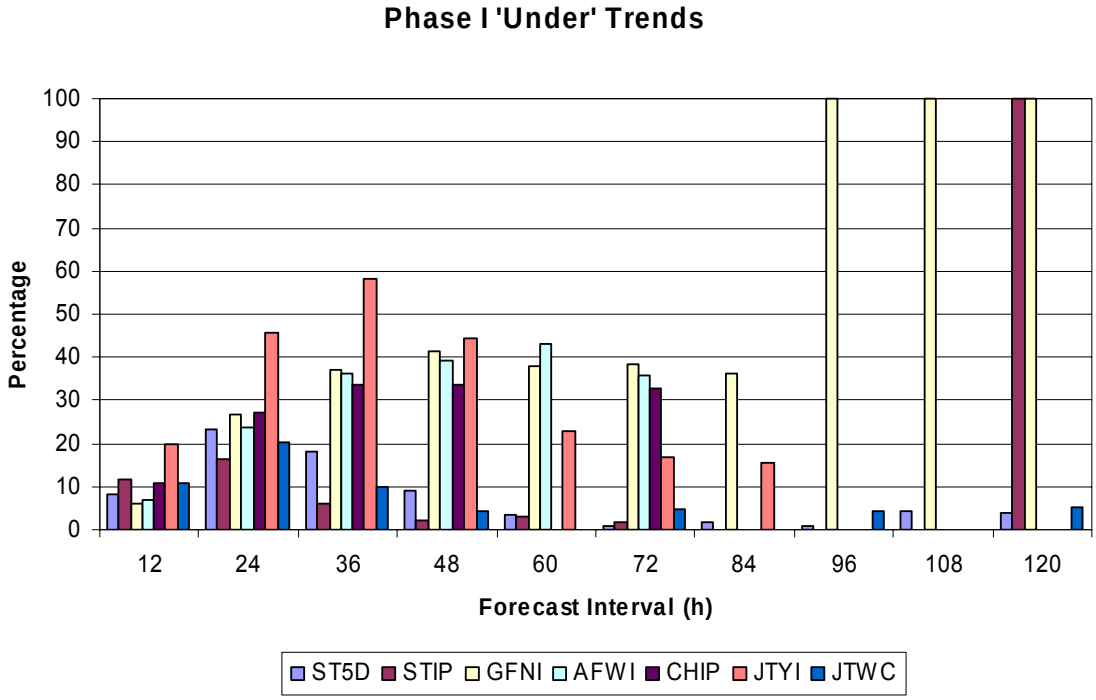

Figure 4.7 Percentage of Under intensity trends as defined in Figure 4.5 during storm formation for the combined 2003 and 2004 seasons. Sample sizes of the verified forecasts range from over 400 at 12 h to less than 30 by the 96-h point. An Under forecast during formation indicates that the magnitude of the forecast intensification rate is less than the magnitude of the actual intensification rate.

3. Over (O) Intensity Trends During Phase I

In general, the percentages of the Over intensity trend forecasts (Fig. 4.8) during Phase I are smaller than for the Under intensity trend forecasts. All of the techniques and JTWC have very small percentages of O forecasts at 12 h. Again using ST5D as the skill measure, the skillful forecasts at 24 h are GFNI, CHIP, and JTYI. Compared to ST5D, the STIP has about twice as many 24-h forecasts that over-intensify the storms during Phase I. Both STIP and JTWC tend to over-intensify the storms at 36 h. By 48 h, all of the other intensity techniques have more skill than ST5D, and the JTWC continues to over-intensify storms in the formation stage. After 48 h, all of the techniques and JTWC have skill relative to ST5D in not over-predicting the intensity trends during formation.

This skill relative to ST5D arises because the ST5D will tend to predict an intensification (formation) too large of a fraction of the time based on climatology.

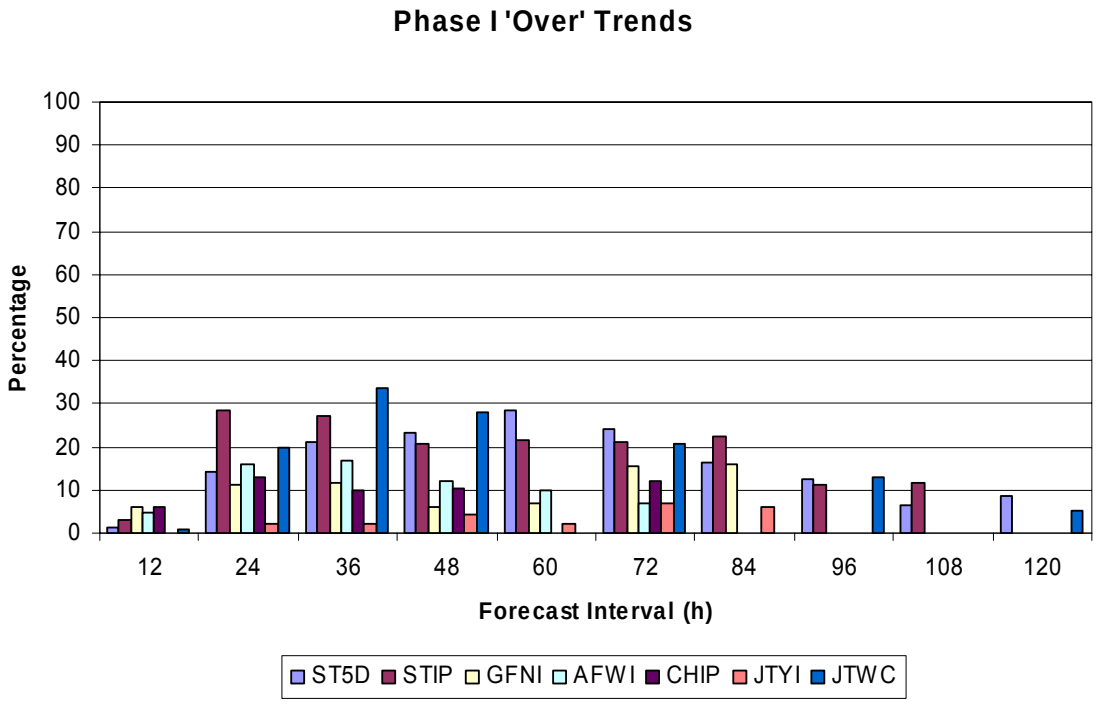

Figure 4.8 Percentage of Over intensity trends as defined in Figure 4.5 during storm formation for the combined 2003 and 2004 seasons. Sample sizes of the verified forecasts range from over 400 at 12 h to less than 30 by the 96-h point. An over-forecast during formation indicates that the magnitude of the forecast intensification rate exceeds the magnitude of the actual intensification rate.

4. Phase I Intensity Trend Summary

Overall, the ST5D and STIP intensity guidance techniques provided the most reliable intensity change trends during Phase I. Both of these techniques generated about 75% G intensity trends averaged over all forecast intervals. The JTWC also had a high G intensity trend percentage (~75%) when averaged over all forecast intervals. This similarity suggests that JTWC follows their best guidance, but because that best guidance is most frequently the ST5D during Phase I, the value-added or skill is small. The GFNI, AFWI, CHIP, and JTYI generally had much lower G percentages between 24 and 48 h, as each technique erred toward an under-forecast of intensity change during formation.

The performance of the official forecasts was also the lowest between 24 and 36 h, when there was a tendency to over-forecast intensification rates.

One possible explanation for this trend in errors during Phase I is based on the average durations for short-lived, average-lived, and long-lived storms that range from 2.5 to 4.25 days (see Chapter III C). That is, the pre-tropical cyclone seedling will have existed about this length of time before it intensified. The various techniques and JTWC forecasters must then predict the timing of this duration before intensification begins. The lowest Good and highest Over intensity change percentages for the 24-h to 48-h forecasts during Phase I suggests that this is the average duration in Phase I before intensification begins. Therefore, one can conclude that this period is longer than the other techniques and JTWC predictions; i.e., they tend to predict formation in the 24-48 h period, which is too soon. Thus, better understanding and forecast techniques for predicting formation is needed.

C. PERFORMANCE DURING INTENSIFICATION PHASE

Whereas intensity guidance was not always available during the formation phase, output from the six techniques was consistently available during the intensification phase. With the larger number of intensity forecasts, several evaluations of Phase II model performance were possible. These evaluations were performed in conjunction with objective analyses to determine qualitatively and quantitatively the intensity forecast trend success during Phase II.

1. Ability to Forecast Phase II Intensity Trends

Referring to the intensity framework (Figures 3.1 and 3.2), any forecast from the time the TC was at or above 35 kt until the time of the first peak intensity was evaluated. All forecast intervals to 120 h were included, so the verification time for the longer forecasts could be in Phase IIa or Phase III. Recall from Fig. 4.5 that the metric was whether the intensity trend of +/- 10 kt over the forecast interval was Good, Under, or Over.

a. Good (G) Intensity Trends During Phase II

For all 12-h forecasts during the time the storm was in Phase II (Fig. 4.9), the ST5D, STIP, and AFWI techniques had percentages of Good (G) approaching 70%, which indicates that these techniques provide intensity change guidance within +/- 10 kt in about 7 of 10 cases. For the 24 h to 84 h forecast intervals, the statistical-dynamical technique STIP had the highest percentage of G forecasts, and this was followed by the ST5D model at 24, 60, 72, and 84 h. Therefore, the STIP-generated forecasts often had some small skill relative to the ST5D during the storm intensification phase. At 36 and 48 h, forecasts produced by the GFNI had the second-highest (out of six techniques) percentage of G trend forecasts, with ~61% and ~64% respectively. Additionally, the GFNI-generated forecasts had the third-highest percentage of G intensity trends at 24, 60, 72, and 84 h.

While the AFWI percentage of Good (G) forecasts from Phase II initial conditions was high at 12 h, the percentage decreased dramatically by 24 h (~44%), probably because of a nearly-steady forecast tendency (systematic error) after initialization. In fact, the AFWI model demonstrated the least success at forecasting intensity trends while the storm was in Phase II, as it had the lowest percentage of G intensity trend forecasts between 24 and 72 h. The percentage of G intensity trends from the JTYI techniques remained near 50% at most forecast intervals. This indicates that nearly one-half of JTYI forecasts produce erroneous intensity change guidance during the intensification phase.

Finally, the JTWC had the highest G intensity trend percentage at all forecast intervals for Phase II. Thus, the official intensity change forecasts added some value to guidance produced by the ST5D or STIP. However, the JTWC percentages of G intensity trend forecasts are only slightly higher than the percentages observed for ST5D and STIP (at all forecast intervals). These similar percentages again indicate that the JTWC relies on those two techniques as their primary guidance for intensity changes during Phase II, but the added value or skill is small.

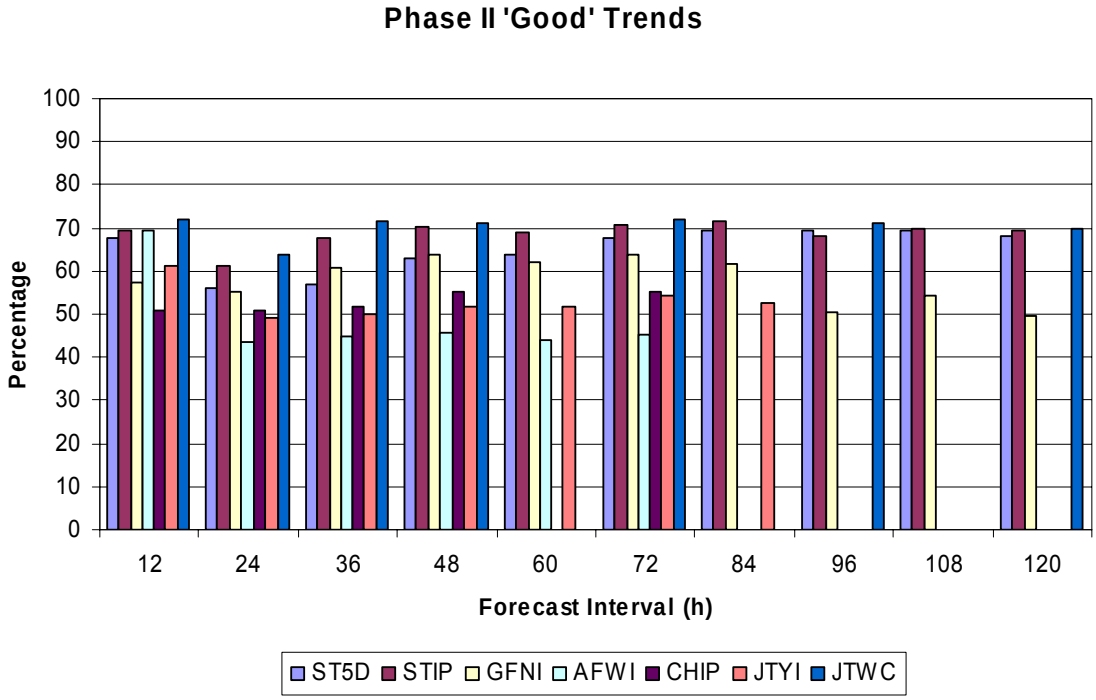

Figure 4.9 Percentage of Good intensity trends as defined in Fig. 4.5 during intensification for the combined 2003 and 2004 seasons. Sample sizes of the verified forecasts range from over 600 at 12 h to less than 200 by the 96-h point. A Good forecast during intensification indicates that the magnitude of the forecast intensification rate is within +/- 10 kt of the actual intensification rate.

b. Under (U) Intensity Trends During Phase II

The percentages of under-forecast intensity trends during Phase II (Fig. 4.10) all exceed 20% for 12-h forecasts. Then the AFWI, CHIP, and JTYI U intensity trends all exceeded 30% at 24 and 36 h, which means that nearly one-third of the intensity change guidance generated by these techniques yielded an under-forecast of intensification. That is, the actual rate of intensification exceeds the predicted rate of intensification by 10 kt or more. This relatively high percentage of JTYI and AFWI under-forecasts is often the result of 'flat' intensity change forecasts. Indeed, the AFWI and JTYI often forecast slow intensification rates out to 24 h, and maintain a nearly-constant intensity forecast from 24 to 48 h. Therefore, the AFWI and JTYI intensity change forecasts are often below the actual intensity change ('undershooting') by the 24-h point (although such a flat

intensity forecast may coincidentally approach the actual intensity value later in the forecast period as the storm begins to decay).

The STIP produced the lowest percentage of U intensity trends at nearly every forecast interval. This was especially evident between 36 and 120 h, when the U percentage remained near 10%. Thus, the STIP rarely generated under-forecasts during storm intensification, and definitely had skill relative to the ST5D in predicting intensity change during Phase II. The GFNI model also produced rather low percentages of U intensity trends that usually remained between 15 and 25%, and had skill relative to the ST5D between 24 h and 60 h. The magnitude of the JTWC under-forecast intensity trends nearly matched the lowest U percentages produced by the STIP technique, which again suggests the reliance that JTWC places on this technique for guidance on intensity changes while the storm is in Phase II.

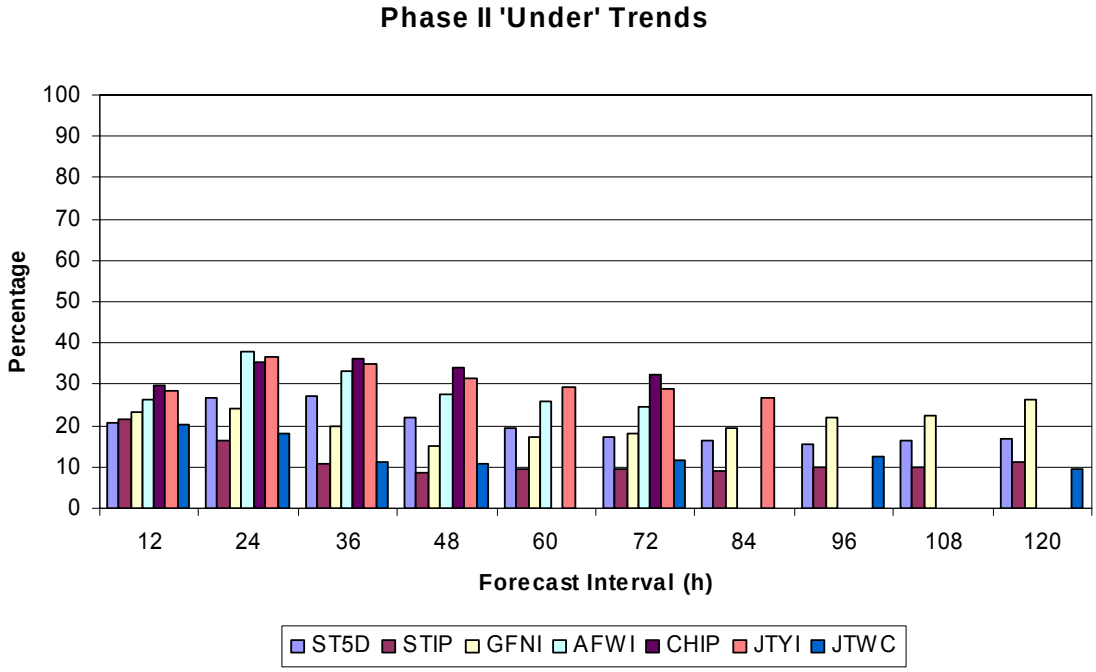

Figure 4.10 Percentage of Under intensity trends as defined in Fig. 4.5 during intensification for the combined 2003 and 2004 seasons. Sample sizes of the verified forecasts range from over 600 at 12 h to less than 200 by the 96-h point. An under-forecast during intensification indicates that the magnitude of the forecast intensification rate is less than the magnitude of the actual intensification rate.

c. Over (O) Intensity Trends During Phase II

Since the Good, Under, and Over percentages must sum to 100, these Over (O) intensity change percentages are the remaining portion of the forecasts. While the AFWI, CHIP, and JTYI techniques generated rather high percentages (> 30% at 24 and 36 h) of under-forecast intensity trends, none of the techniques produced over-forecasts greater than 30% of the time. In fact, only the AFWI had the 30% O value at 60 and 72 h. Otherwise, the over-forecast percentages from all of the remaining techniques never exceeded ~22%. That is, less than one-in-four intensity change forecasts were over-forecasts of intensity during Phase II. In fact, the intensity change guidance at many forecast intervals only generated O intensity trends ~15% of the time. Using ST5D as a skill measure (smaller percentages of U forecasts): 12 h, STIP, AFWI, JTYI, and JTWC had skill; 24 h-36 h, CHIP and JTYI; 48 h and 72 h: CHIP; and after 72 h none of the techniques had skill relative to ST5D.

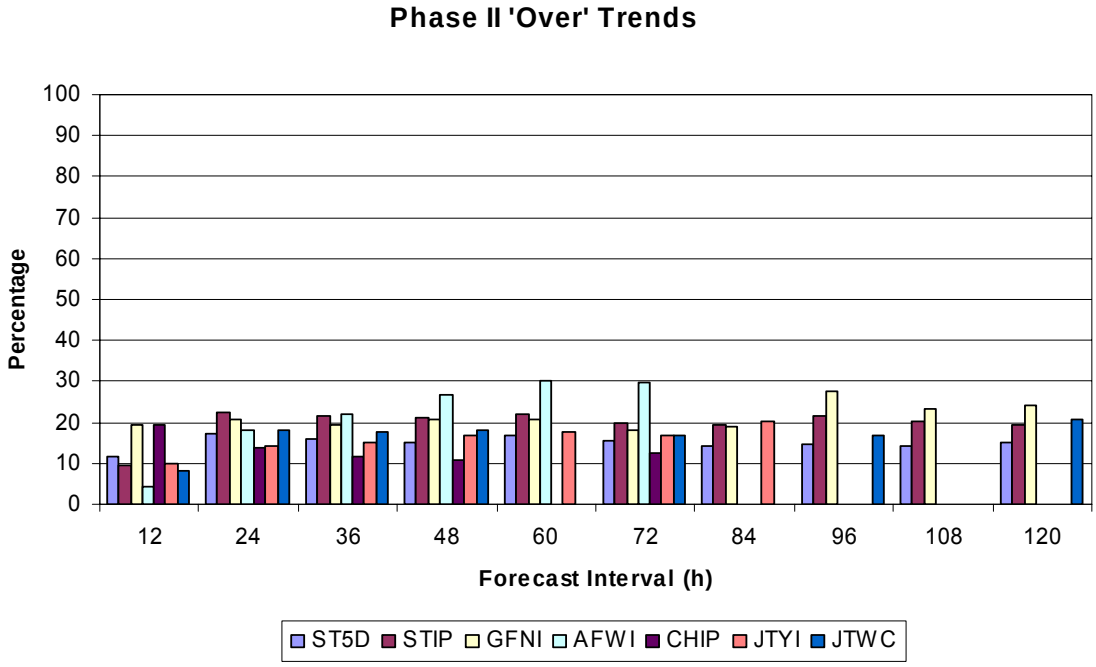

Figure 4.11 Percentage of Over intensity trends as defined in Fig. 4.5 during intensification for the combined 2003 and 2004 seasons. Sample sizes of the verified forecasts range from over 600 at 12 h to less than 200 by the 96-h point. An over-forecast during intensification indicates that the magnitude of the forecast intensification rate is greater than the magnitude of the actual intensification rate.

d. Phase II Intensity Trend Summary

This evaluation of the capability of the various techniques to predict intensity trends according to the categories in Fig. 4.5 indicated that the statistical-dynamical STIP, followed by the statistical ST5D and the GFNI, produced the most reliable intensity change guidance during the intensification phase. Indeed, the STIP had G intensity trends 68.5% of the time when averaged over all of the forecast intervals. Additionally, nearly two-thirds of all ST5D and GFNI forecasts had G intensity trends when averaged over all of the forecast intervals. With few exceptions, the three remaining dynamic techniques (AFWI, CHIP, and JTYI) demonstrated no skill relative to the STIP or ST5D. In fact, the AFWI had the lowest G intensity trend percentage relative to all of the intensity guidance techniques during intensification. Conversely, the JTWC forecasts outperformed guidance from every technique, as nearly three-quarters of the official forecasts produced G intensity change trends.

2. Errors Verifying at Peak Intensity

While the above intensity trend analysis indicated general technique performance during Phase II, it is also of interest whether the techniques are capable of predicting the peak intensity at the end of Phase II. Therefore, average intensity errors for the various techniques for the forecasts preceding the time of maximum intensity were computed to quantify intensity guidance reliability. For each of the 59 storms in the 2003-2004 database, the date-time-group (DTG) of the (first) peak intensity was determined for the series of intensity predictions verifying at this time of peak intensity (-120 h, -96 h,…-12 h, 0 h). The averages of all predictions of peak intensity minus the actual intensity for the various forecast intervals that could be verified are shown in Fig. 4.12. In this case, a non-homogeneous sample is allowed to maximize the sample sizes.

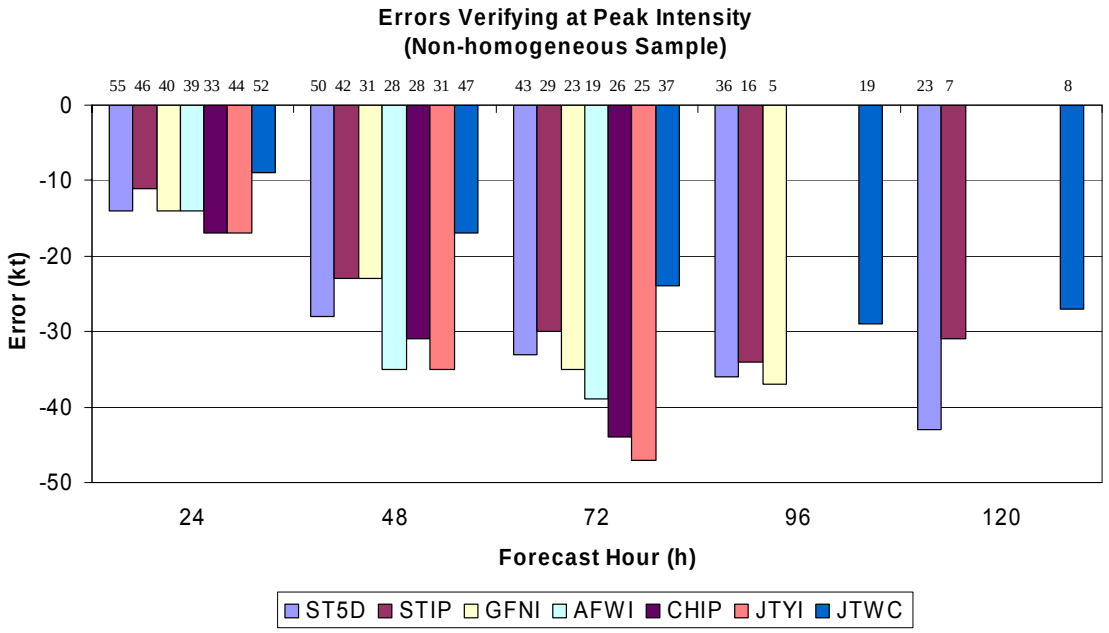

Figure 4.12 Intensity forecast errors (kt) at the date/time of the first peak intensity of each of the 59 TCs during 2003 and 2004 by the various techniques and the JTWC (see insert). The small numbers above each bar indicate the number of cases.

Even only 24 h before the time of peak intensity, all of the intensity prediction techniques and JTWC forecasts have an average error of -9 kt or more (under-forecasts). Using ST5D as a skill measure, only the STIP has skill among the five other intensity prediction aids at 24 h, and yet JTWC forecasters are able to use this guidance (and other information) to add value. These features continue at 48 h, except the tendency for under-forecasting of the peak intensity is even stronger, and only the STIP and GFNI have skill relative to ST5D. The JTWC forecasts at 48 h prior to peak intensity are on average about 17 kt too low. The same trend toward under-forecasting of the peak intensity exists at 72 h. Only the STIP has a little skill relative to the ST5D, and the JTWC continues to have the most skill, although they under-forecast the peak intensity by nearly 24 kt on average. Beyond 72 h, the number of available intensity techniques decreases, and the sample sizes also decrease because few forecasts of peak intensity 96 h or 120 h into the future may be validated. Thus, not too much confidence can be placed in these calculations. Those 120-h JTWC forecasts of peak intensity are under-forecasts by about 28 kt on average.

3. Rapid Intensification

Overlays of storm intensity and the intensity forecasts (e.g., Fig. 1.1) displayed for all storms during the 2003 and 2004 seasons revealed many of the techniques missed periods of large intensity change. That is, for most of the techniques, the predicted rates of intensification were often less than the actual rate of intensification. As a result, many of the techniques under-forecast storm intensity during Phase II.

a. Definition

The majority of intensity forecast techniques could not predict intensification rates equal to or exceeding 30 kt/day. That is, an 'upper bound' on predicted intensification rates for most of the techniques seems to be at or below 30 kt/day. Therefore, 'rapid intensification' (as applied to this thesis) refers to a forecast or observed intensity increase equal to or exceeding 30 kt/day. As a comparison, Holliday and Thompson (1979) defined western North Pacific rapid intensification as a 24-h central pressure fall equal to or greater than 42 mb (top 25% of all intensification cases).

Based on estimates of model-predicted intensity change from the storm overlays, the majority of the techniques, a majority of the time, predicted intensity increases between 10 and 20 kt per day, which may be regarded as a 'typical' rate of intensification. The middle ground between typical intensification and 'rapid' intensification would then be considered an 'above-average' rate of intensification. Finally, intensity change increases less than 10 kt/day are considered 'slow'.

b. Subjective Evaluation

Having defined the threshold for rapid intensification, all intensity forecast techniques (and the JTWC intensity change forecast) were evaluated as to their capability to predict rapid intensification. Predictions by each technique were examined for all times +/- 24 h relative to the actual time. If any of these predictions matched (or exceeded) the rapid intensification thresholds, it was regarded as a hit. Conversely, if none of the predictions within +/- 24 h matched or exceeded the threshold, then the event was recorded as being missed.

Overall, 28 cases of rapid intensification were observed during 2003/2004 (Figure 4.13). The ST5D, STIP, and AFWI did not provide any guidance as to rapid intensification events with zero, one, and zero (respectively) correct predictions of the 28 observed periods of rapid intensification. Likewise, the three and six successful predictions (within +/- 24 h of the actual event) by the JTYI and CHIP models are not really useful guidance. Whereas the GFDI model did predict 10 of the 28 rapid intensification events, it has a tendency to over-predict intensification. Based on this guidance, JTWC was able to forecast only 6 of the 28 events within +/- 24 h of the time of actual rapid intensification.

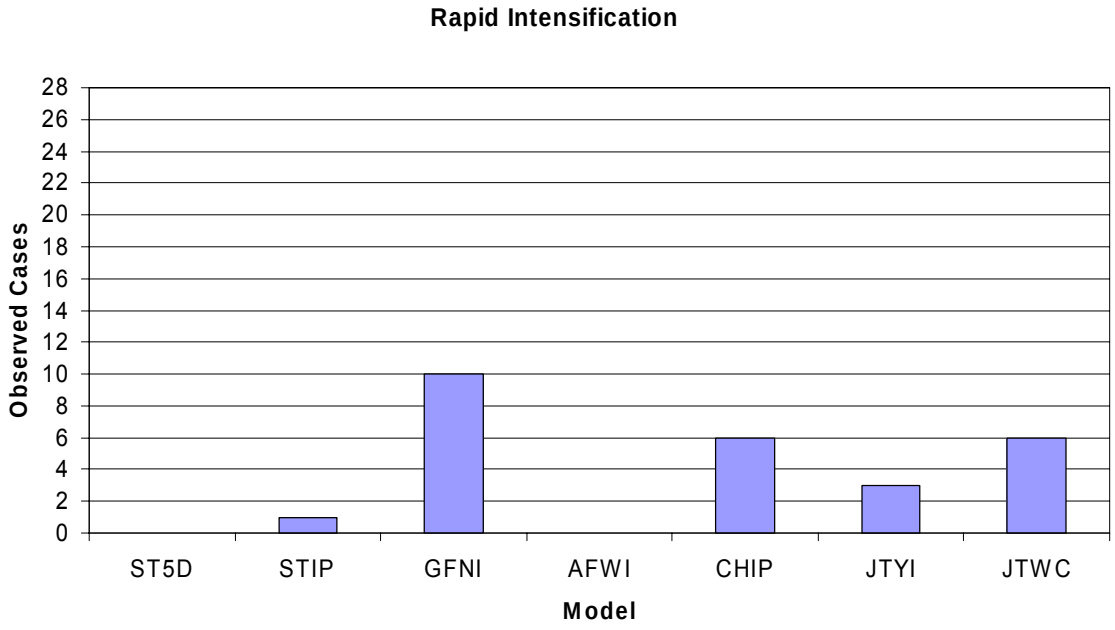

Figure 4.13 The number of cases that a forecast technique or JTWC adequately predicted an observed period of rapid intensification.

4. Distribution of Intensification Rates During Phase II

A skillful intensity guidance technique should have two properties: (i) The mean of the intensity change predictions should be equal to the mean of the observed intensity changes, i.e., the technique should not have a bias; and (ii) The distribution of intensity changes predicted by the technique should be similar to the distribution of observed intensity changes. The analyses in the preceding subsections of the intensity changes at various time intervals during Phase II

indicate biases exist and that the intensity techniques do not provide good guidance for rapid intensification events.

In this section, the observed distribution of 48-h intensity changes from the beginning of Phase II (initial intensity of 35 kt) will be compared with the predicted distributions by the various techniques. Although other time intervals were examined, the first 48 h of Phase II may best distinguish between the durations and characteristic intensification rates. Recall from Chapter III.C that the average durations in Phase II are 1.5, 3.5, and 7.0 days for short-lived, average-lived, and long-lived storms. So, the 48-h period for an average-lived storm is half the expected intensification period, and on average the intensification will be over for a short-lived storm. The results of this evaluation are displayed in Figure 4.14 (a-g).

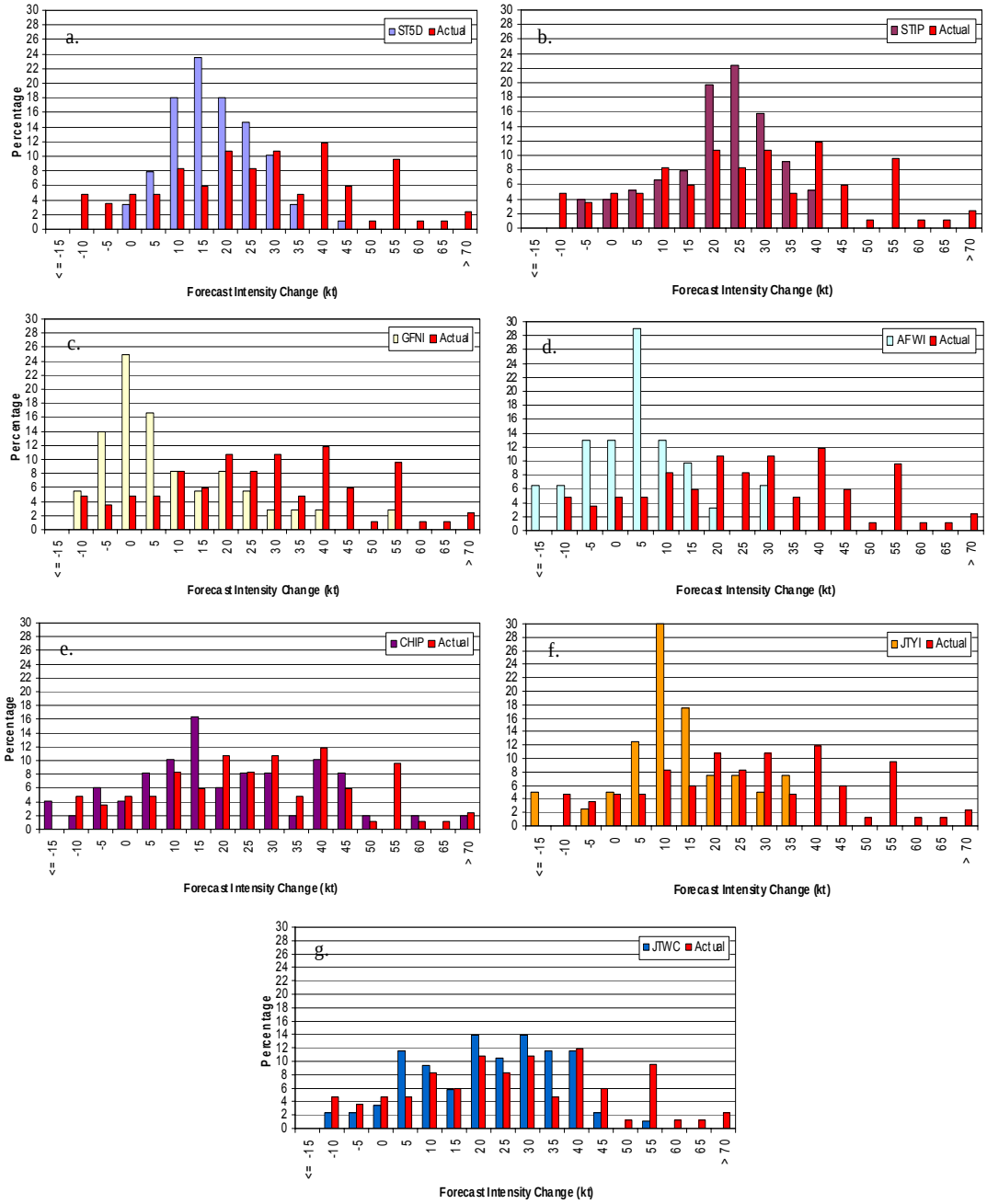

Figure 4.14 (a-g) Phase II graphs of actual 48-h intensity change distribution plotted with the 48-h intensity change distributions of all the various techniques and JTWC.

The observed 48-h intensity change distribution, which is repeated in each panel in Fig. 14, ranged from –10 kt (indicates some cases of decay after 48 h since the beginning of Phase II) to greater than 70 kt. The highest percentages of 48-h intensity change for this small sample were at 20, 30, and 40 kt. Almost

one-half of all 48-h intensity changes ranged between 20 and 40 kt, and just over three-quarters of the observed changes were between 10 and 55 kt.

The histogram of observed intensity changes are compared with the climatology and persistence technique (ST5D) in Figure 4.14a. The ST5D rarely produced rates lower than 5 kt or higher than 35 kt, and thus has a relatively narrow intensity change distribution relative to the observed distribution. Nearly 60% of the ST5D 48-h intensity change forecasts had intensification rates between 10 and 20 kt, and 92% of these forecasts remained between 5 and 30 kt. Thus, not only does the ST5D have a bias toward lower 48-h intensity changes, it does not provide guidance on the larger observed intensity changes in Phase II.

Similarly, nearly 58% of the statistical-dynamical technique STIP intensification rates from the beginning of Phase II were 20-30 kt/48 h (Fig. 4.14b). While the distribution of STIP predictions in this small sample ranged from –5 kt/48 h to 40 kt/48 h, the STIP technique infrequently generated intensification rates less than 10 kt or greater than 35 kt. Thus, the STIP technique will over-intensify the weak intensifiers (or the 48-h decay cases) and under-intensify the more rapid intensifiers at the beginning of Phase II.

Whereas the ST5D and STIP techniques had positively skewed peaks in their distributions, the GFNI peak 48-h intensity change (Fig. 4.14c) was centered about 0 kt (i.e., no change), which indicates the GFNI model has little skill in the early intensification of Phase II. Indeed, ~55% of GFNI intensification rates were between –5 kt and 5 kt. Sometimes the GFNI model will provide an indication of intensification, and nearly 28% of the GFNI intensification rates ranged from 10 kt/48 h to 25 kt/48 h. However, the GFNI model does not predict the typical distribution of intensification rates or the largest values.

The peak of the AFWI distribution (Fig. 4.14d) was strongly centered on 5 kt, with over 75% of the 48-h intensity change forecasts having rates between –5 kt and 15 kt. As will be shown in Fig. 4.20 below, many AFWI forecasts had a flat intensity profile—especially beyond 24 h. The very different observed and

AFWI distributions in Fig. 4.14d clearly indicate the AFWI will not provide useful guidance in the early intensification stage of Phase II.

The coupled ocean-atmosphere model (CHIP) 48-h intensification rates (Fig. 4.14e) varied over a large range, as this technique produced 48-h intensity change forecasts between –15 kt and 90 kt. Unfortunately, the CHIP forecasts for seven of the 22 most intense (110 kt or greater) storms were unavailable. The peak of the CHIP distribution was centered on 15 kt, with just over 60% of the values between 0 kt and 30 kt. Of the techniques in this study, CHIP produced the largest intensification rates, as two outlier forecasts reached 90 kt/48 h. However, the CHIP model also had some large 48-h decay values that did not verify.

The JTYI model was similar to the GFNI and STIP models tendency to generate conservative 48-h intensity change forecasts in Phase II (Fig. 4.14f). The peak of the JTYI distribution was strongly centered on 10 kt, and 60% of the forecasts ranged between 5 and 15 kt. Only 25% of the JTYI-generated forecasts ranged from 20 to 35 kt, and no forecasts exceeded 35 kt/48 h. Therefore, the JTYI model also does not provide good guidance as to the early intensification rates in Phase II.

Finally, the JTWC 48-h intensity change distribution (Fig. 4.14g) is similar to the observed distribution in the range from –10 kt to 40 kt. In fact, about ~88% of the JTWC 48-h intensity change forecasts were between 5 and 40 kt. However, the JTWC rarely forecast intensification rates greater than 40 kt/48 h, and thus missed the highest intensification rates (top ~20%). Nonetheless, the JTWC produced a reasonable distribution of intensity change forecasts, which is encouraging given the rather limited range of intensity change distributions of the intensity guidance techniques that they had available.

An additional representation of the 48-h intensity change distributions is given in Figure 4.14h. This summary clearly indicates that the averages of all the predicted intensity change distributions remain below the average of the actual distribution, which reaffirms that all of the intensity guidance techniques and the

JTWC forecasts have a bias of under-estimating the rates of intensification during Phase II. The too-low intensity change bias of the dynamical models (GFNI, AFWI, and JTYI) is particularly noteworthy. Except for the CHIP model, all of the intensity guidance techniques and the JTWC have too narrow ranges of intensity change forecasts. In particular, none of the techniques or JTWC has the proper number of rapid intensification forecasts during the early intensification stage of Phase II. The too-low bias and too-small range of intensity changes for the dynamical models (GFNI, AFWI, and JTYI) leads to almost no overlap with the actual intensity change distribution. Because much of the guidance does not even fall within the actual intensity change distributions, it is clear that improvements in the intensity forecast guidance are needed, and particularly for rapid intensification events.

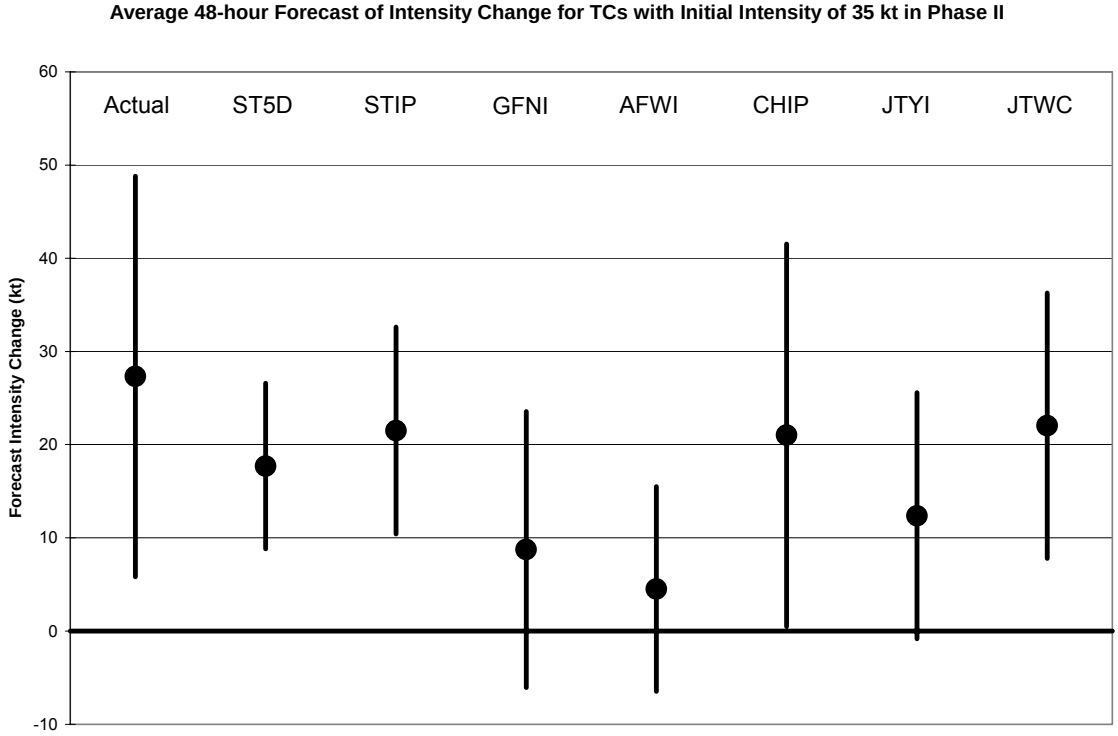

Figure 4.14h Stick diagram of 48 h intensity change distributions for 59 TCs during the 2003-2004 seasons with an initial intensity of 35 kt. The diagram includes the actual intensity change distribution and the distributions of all the various techniques. The heavy dot represents the mean intensity change value (kt) while the length of the sticks represents plus or minus one standard deviation about the mean.

D. PERFORMANCE DURING DECAY/REINTENSIFICATION CYCLES

Secondary peaks, whether due to internal storm processes or caused by environmental factors, correspond to Phase IIa of the intensity framework (see Fig. 3.2). The capability of the techniques to predict changes in Phase IIa was examined with an assessment of 'secondary peaks.' For this thesis, a secondary peak is defined as a reintensification of 10 kt or greater following an initial decay cycle of 10 kt or greater (e.g., eyewall replacement cycles).

Predictions by each technique were examined for all times +/- 24 h relative to the actual time of secondary peaks. If any of these predictions reasonably matched the phase (timing) and magnitude of intensity oscillations, it was regarded as a 'hit'. Conversely, if none of the predictions within +/- 24 h

reasonably matched the phase and magnitude of intensity oscillations, then the event was recorded as being missed.

Overall, 12 cases of secondary peaks were observed during the 2003 and 2004 seasons (Figure 4.15). None of the techniques captured more than 25% of observed secondary peaks. The climatology and persistence technique ST5D registered zero hits, and the statistical-dynamical STIP had only one. One could surmise that the ST5D and STIP tend to expect only one intensity peak (especially when the intensity oscillations occur over a relatively short time period). While the dynamic model techniques (GFNI, AFWI, CHIP, and JTYI) were an improvement over the ST5D and STIP, they still missed almost all of the secondary peaks. The question is whether even with increased spatial/temporal resolution and better model physics will these techniques predict eyewall replacement cycles or other internal processes that lead to secondary peaks. Qualitatively, the GFNI forecasts produced intensity oscillations more often than the other dynamic models, but the timing and magnitude of the oscillations often did not meet the thresholds required to count as a 'hit.' Despite a dearth of skillful guidance, the JTWC outperformed all of the techniques by recording five 'hits' within +/- 24 h of the observed secondary peaks.

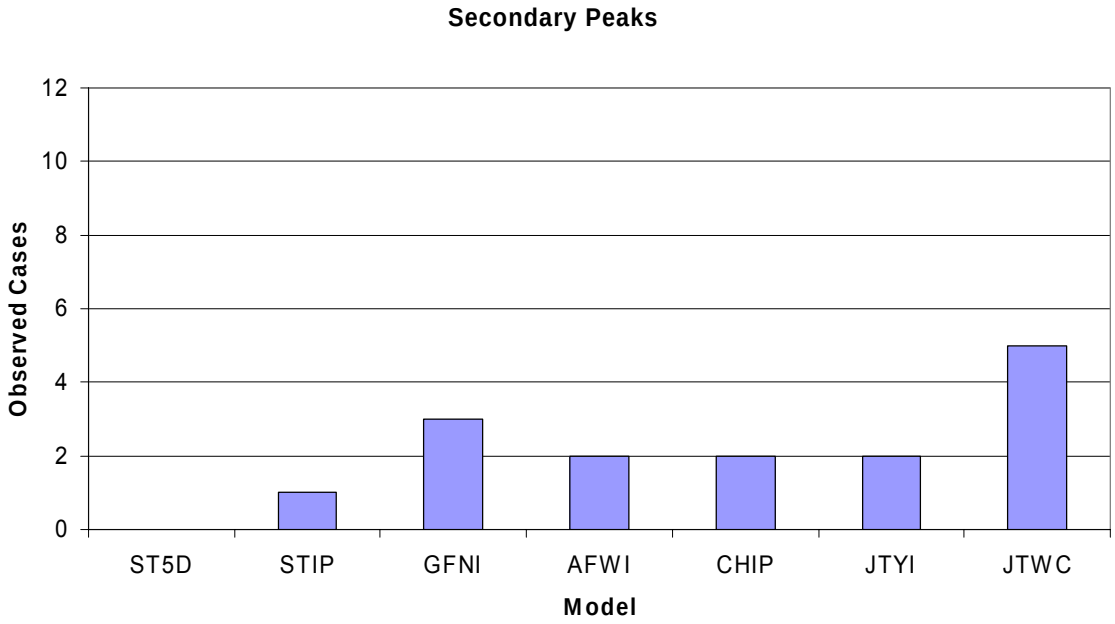

Figure 4.15 Number of cases (out of 12) a forecast technique adequately predicted an observed secondary intensity peak.

E. PERFORMANCE DURING DECAY PHASE

1. Ability to Forecast Phase III Intensity Trends

Similar to intensity trend analysis for Phase II, all of intensity technique predictions during Phase III were examined for the trend thresholds in Fig. 4.5 (from peak intensity through final decay, or end of best track data). If the storm had a decay/reintensification cycle (i.e., Phase IIa), then the primary decay phase (Phase III) started at the DTG of the secondary (or last) intensity peak and continued until final decay.

a. Good (G) Intensity Trends During Phase III

All of the intensity guidance techniques and the JTWC forecasts had rather high percentages of G intensity trend forecasts (Fig. 4.16) at 12 h (> 60%). However, the ST5D (or the STIP) had the highest percentage of G intensity trends out to 84 h, and consequently, none of the other techniques had skill relative to the these two models. In fact, the ST5D performance actually improved between 48 and 72 h, when this technique had G intensity trends over 80% of the time. The percentage of GFNI G intensity trends gradually decreased between 12 h (~68%) and 48 h (~52%), and then increased from 48 h to 84 h

(~75%). That is, the GFNI performed best at the early and later forecast intervals during the decay phase.

While the AFWI technique G intensity trends were above 60% at 12 h, the percentage of G intensity trends dramatically declined between 24 and 48 h, and reached a low of ~23% G intensity trends for the 48-h forecasts. Indeed, the percentage of G trends remained under 40% from 24 h to 72 h, which indicates that the majority of AFWI forecasts are poor during the decay phase. Good intensity trend guidance from the CHIP technique remained above 60% at all forecast intervals, and these G trends actually increased between 36 h (~65%) and 72 h (~84%). Although the CHIP outperformed four of the other intensity change guidance techniques at 72 h, it was not better than ST5D and thus can not be said to have skill. The percentage of G intensity trends for JTYI decreased between 12 h (~69%) and 48 h (~44%) and then increased from 48 h to 84 h (~68%). Thus, the majority of JTYI forecasts are poor during the middle of the forecast interval. For JTWC, the percentage of G intensity trends remained above ~69% at all forecast intervals, but only had skill relative to both the ST5D and STIP at 12 h and 96 h. However, sample sizes of verifying forecasts decrease dramatically beyond 84 h so that the higher-than-expected G intensity trend percentages between 96 and 120 h are not likely to be significant.

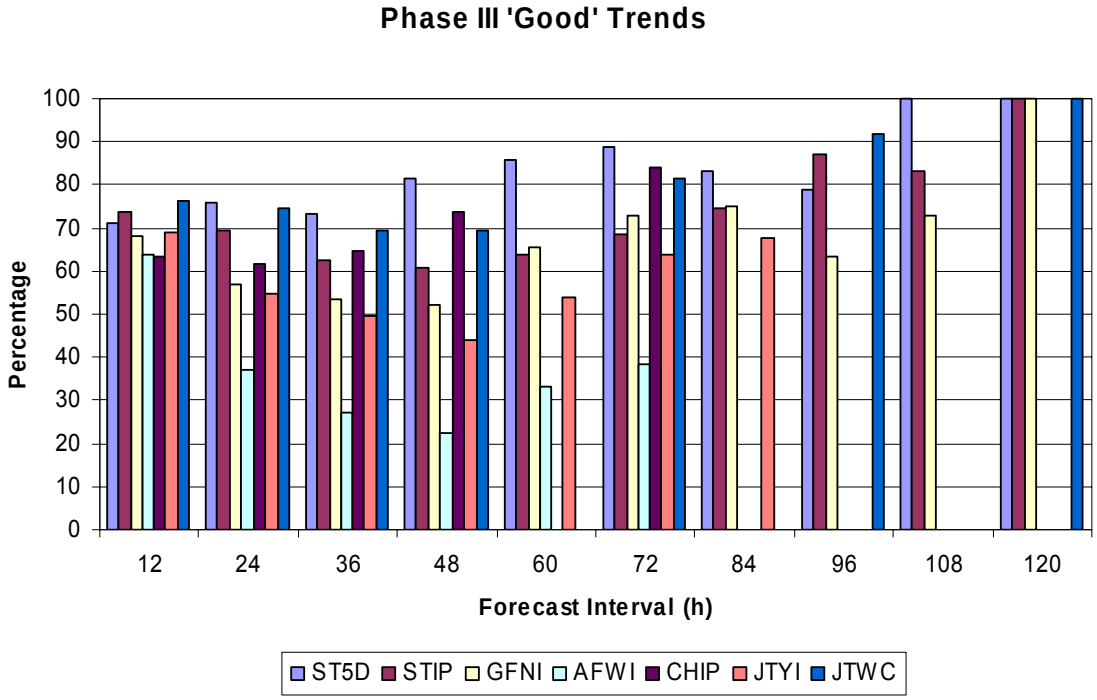

Figure 4.16 Percentage of Good intensity trends during the decay Phase III for the combined 2003 and 2004 seasons. Sample sizes of the verified forecasts range from 400 at 12 h to less than 30 by the 96-h point. A Good forecast during decay indicates that the magnitude of the forecast decay rate is within +/- 10 kt of the actual decay rate.

b. Under (U) Intensity Trends During Phase III

The percentage of U intensity trends during the decay Phase III (Fig. 4.17) remained very low (almost all below 17%) at all forecast intervals. These low U intensity trend percentages indicate that nearly all of the techniques (at all forecast intervals) do not generate excessive rates of decay. The one exception is the CHIP technique, which produced a ~26% U intensity trend at 12 h. That is, nearly one-fourth of all CHIP forecasts verifying at the 12-h interval had intensity decreases that exceeded the actual rate of decay. Except for CHIP, the other techniques and JTWC have skill relative to ST5D in not having too many cases of overly rapid decay.

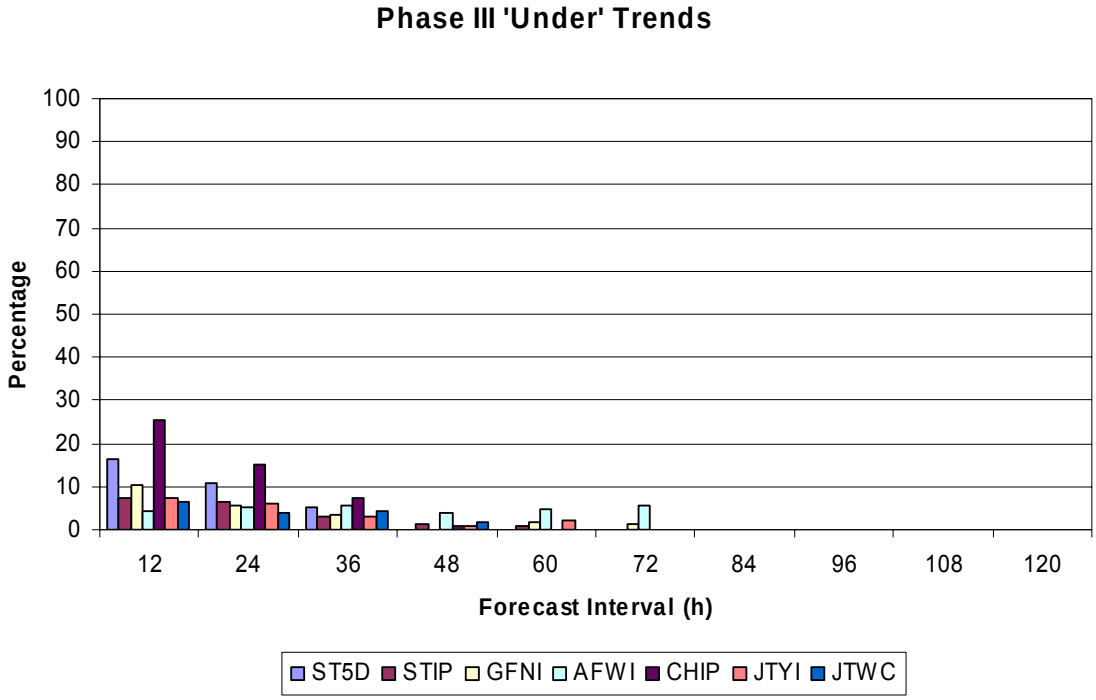

Figure 4.17 Percentage of Under intensity trends during the decay Phase III for the combined 2003 and 2004 seasons. Sample sizes of the verified forecasts range from over 400 at 12 h to less than 30 by the 96-h point. An under-forecast during decay indicates that the magnitude of the forecast decay rate is greater than the magnitude of the actual decay rate.

c. Over (O) Intensity Trends During Phase III

For most of the techniques, the percentage of O intensity trends (Fig. 4.18) increased between 12 and 48 h, and then decreased between 48 and 84 h. Thus, the cases of over-forecasts were highest during the middle (~48 h) of the forecast interval for a majority of the techniques. The large percentage of Over forecast trends means that many intensity forecasts were greater than the observed verifying intensity, implying that the techniques do not decay a TC fast enough. The AFWI technique produced the most dramatic examples of over-forecasts during Phase III, which indicates especially poor performance during the decay cycle. For example, over 60% of all forecasts at 36, 48, and 60 h provided erroneous guidance on the rate of decay, and the performance was not much better for the 24- and 72-h forecasts. Similarly, over one-third of all JTYI forecasts between 24 and 60 h were over-forecasts during the decay phase, with

the worst performance for the 48-h forecasts (~55% U). As mentioned earlier in the Phase II trend analysis, the AFWI intensity guidance (and to a lesser extent the JTYI) was often for flat intensity trends. Thus, these O intensity trends indicate that the AFWI and JTYI cannot predict the actual rate of decay (i.e., the observed rate of decay remains at least 10 kt above the forecast rate of decay).

The percentages of GFDI over-forecasts between 12 and 72 h were also rather large, especially at 24 h (~38%), 36 h (~43%), and 48 h (~47%). Again the GFNI-generated forecasts performed the worst during the middle of the forecast interval. The STIP had rather large percentages of O intensity trends between 36 and 60 h, when greater than one-third of all STIP forecasts were over-forecasts during the decay phase. The CHIP technique had the second-lowest percentages of over-forecasts during the decay phase, which indicates that this model often produces decay rates that are greater in magnitude than the other techniques.

In contrast, the ST5D technique had the lowest percentage of O intensity trends at all forecast intervals, as the majority of ST5D percentages remained well below 20%. Therefore, none of the intensity techniques had skill relative to the ST5D during the decay Phase III in terms of the intensity trend analysis as in Fig. 4.5. Lastly, Phase III forecasts produced by the JTWC had rather low percentages of over-forecasts when compared to most dynamic techniques. However, the official forecasts also did not have skill in relation to the ST5D during the decay phase.

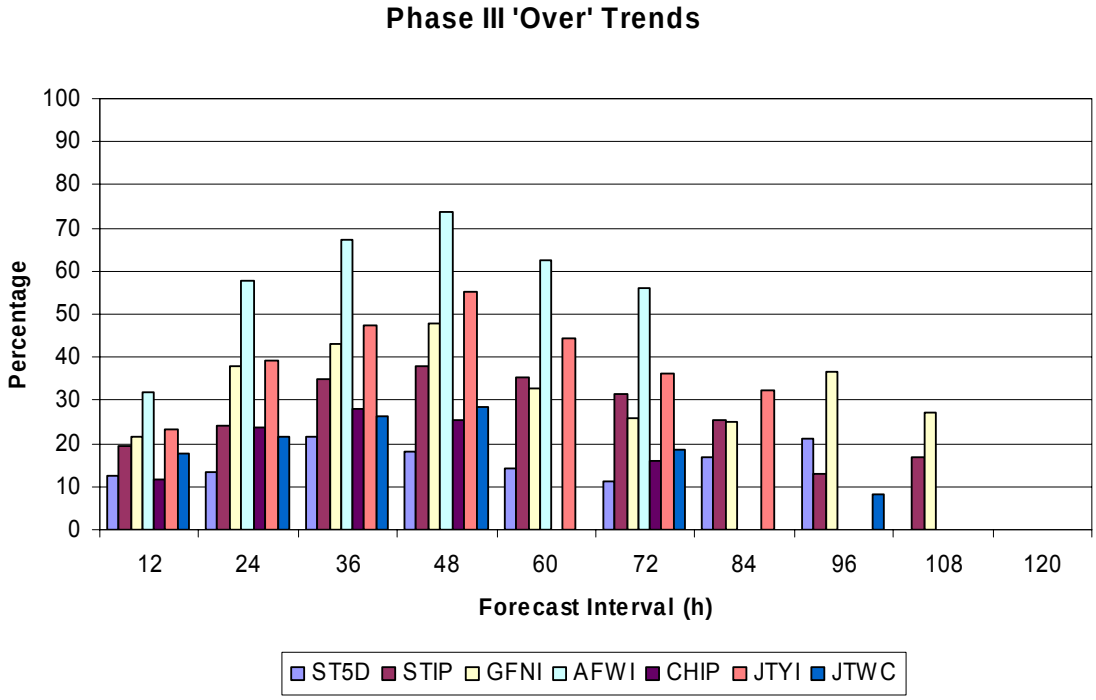

Figure 4.18 Percentage of Over intensity trends during the decay Phase III for the combined 2003 and 2004 seasons. Sample sizes of the verified forecasts range from 400 at 12 h to less than 30 by the 96-h point. An over-forecast during decay indicates that the magnitude of the forecast decay rate is less than the magnitude of the actual decay rate.

d. Phase III Intensity Trend Summary

This evaluation of the capability of the various techniques to predict intensity trends (according to the categories in Fig. 4.5) is rather discouraging in that the ST5D produced the most reliable intensity change guidance during the decay phase. As a result, none of the other techniques added value relative to the no-skill ST5D. While the CHIP did not produce accurate forecasts during intensification, this technique was second only to the ST5D in performance during decay. The STIP had the third-best performance in Phase III. Surprisingly, The AFWI, JTYI, and GFNI performed the worst during decay, as these techniques consistently had decay rates less than the actual rate of decay. While the JTWC forecasts added value relative to many of the techniques, the official forecasts infrequently added value relative to the ST5D.

2. Intensity Forecasts Verifying at the 45-kt Decay Point

A 'damaging wind' threshold of 50 kt is often set at military bases. Just as forecasting onset of 50 kt winds is important, the time at which the winds decay below 50 kt (say to 45 kt) is also a relevant threshold (e.g., when base operations may resume and recovery operations begin). Therefore, it is of interest whether the intensity techniques can provide accurate guidance of the decay to 45 kt.

In this evaluation, the DTG in each of the 26 storms that had decay down to 45 kt (or the DTG immediately following 45 kt if the storm decay passed through the 45-kt point) was extracted. Then all forecasts (i.e., -120 h, -96 h,… -12 h) that could be verified at the DTG of 45 kt decay point were collected. The difference between the predicted intensity and the actual intensity (45 kt) is defined as the error. Decays to other intensities were examined, but the 45 kt value yielded the largest number of cases, in part because many storms are not tracked during an extratropical transition phase or after landfall (where intensity abruptly ends above 45 kt). Weak storms (50 kt or less) were also excluded in this analysis, primarily because they did not exhibit a well-defined decay phase (or were already at the decay threshold).

Since the average 24-h forecast errors verifying at 45 kt in the decay phase (Fig. 4.19) are positive, a tendency exists for the intensity prediction techniques and the JTWC to under-forecast the decay rate on average. Using ST5D as a skill metric, only the CHIP and JTWC have some skill in this aspect of the decay phase. The other techniques (STIP, GFNI, AFWI, and JTYI) all have no skill compared to the ST5D. Given the typical performance of the AFWI model (Fig. 4.20) for the decay phase, it is not surprising that this model already by 24 h has 25 kt under-forecasts of the decay rate on average.

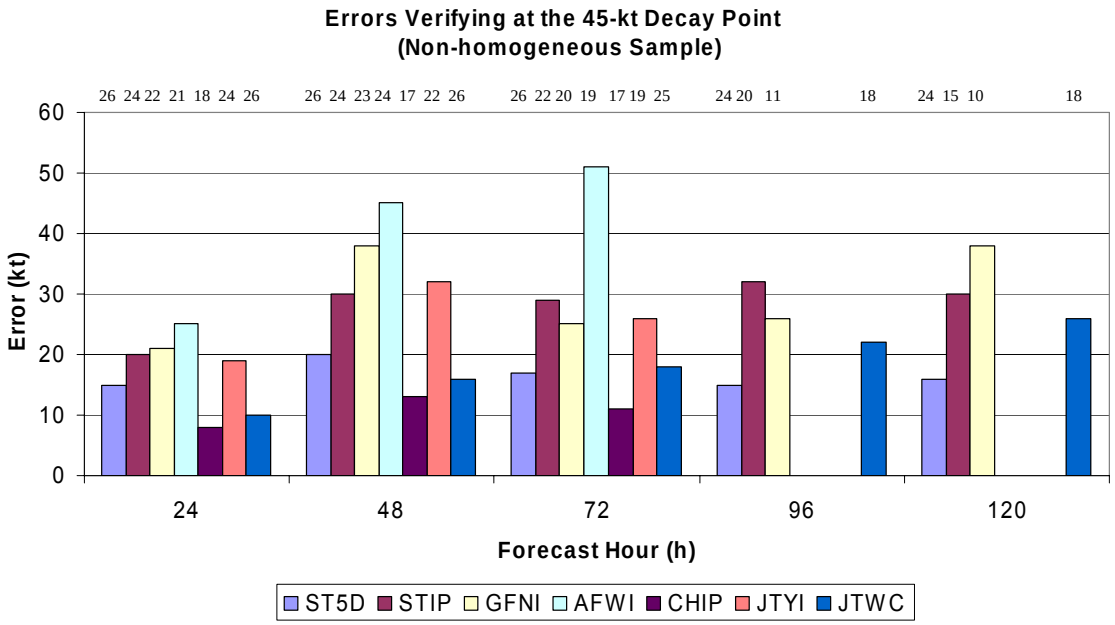

Figure 4.19 Observed intensity errors (kt) using the date/time of decay down to 45 kt as a verification point. The small numbers above each bar represent the number of cases. Note again that a non-homogeneous sample is allowed to maximize the members of cases for analysis.

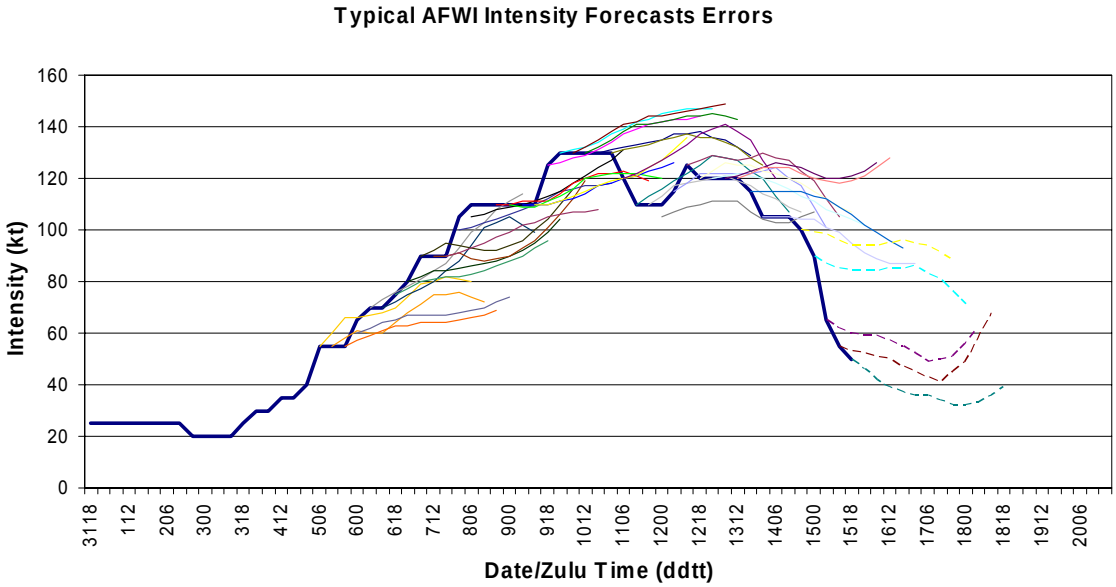

Figure 4.20 Common intensity change forecast errors for the AFWI. The heavy blue line represents observed storm intensity, while the colored lines indicate individual model forecasts every 6 h. Forecasts every 6 h during Phase III fail to capture the actual decay rate.

Similar error characteristics as at 24 h exist for the average 48-h forecasts verifying at the 45-kt decay point (Fig. 4.19). That is, only CHIP and JTWC have skill relative to ST5D, although both under-forecast the decay rate by about 15 kt/48 h on average. It is somewhat surprising that the STIP under-forecasts the decay rate (to the 45 kt threshold) by 30 kt/48 h on average, which suggests that the sample examined here has very different characteristics from the developmental sample for STIP. Whereas the dynamical models (GFNI and AFWI) have even larger under-forecasts of the decay rate to the 45 kt threshold, the JTYI has about the same magnitude error as STIP (but is not skillful relative to ST5D).

Only the CHIP model has skill at 72 h in predicting the decay to 45 kt. Since STIP, GFNI, and JTYI errors at 72 h are smaller than at 48 h, this is probably indicative of too small sample sizes of forecasts that can be validated. Similarly, the sample sizes for 96 h and 120 h forecasts of decay to 45 kt are too small for conclusive results. For those JTWC forecasts that are available, they do not have skill relative to ST5D.

3. Rapid Decay

a. Definition

In the case of rapid decay, the stronger tropical cyclones (i.e., storms reaching typhoon intensity) were chosen due to data availability, since many of these storms have at least one period of rapid decay. The majority of intensity change techniques could not predict observed decay rates equal to or exceeding 30 kt/day. Therefore, the term ‘rapid decay’ (as applied to this thesis) refers to forecast or observed intensity decreases equal to or exceeding 30 kt/day.

b. Subjective Evaluation

The analysis of rapid decay followed the same procedure as the analyses of rapid intensification and secondary peaks. The technique was scored as a ‘hit’ if any forecast by the technique within +/- 24 h of the observed event predicted the rapid decay threshold defined above. Just as for rapid

intensification, 28 storms had at least one period of rapid decay, and of these 28 storms, 23 were typhoons that also had a period of rapid intensification.

With the exception of the AFWI, the techniques were able to forecast more cases of rapid decay than cases of rapid intensification (Fig. 4.21). Of all the techniques, the CHIP captured the greatest number of rapid decay cases with 15. The GFNI was next, as it predicted a period of rapid decay in 12 cases. Unlike the rapid intensification phase when an intensification rate 'upper bound' seemed to apply for ST5D and STIP, the rate of intensity decreases during Phase III often exceeded 30 kt/day. In fact, some ST5D forecasts generated decay rates of about 35 kt/day for the strongest typhoons (110 kt or greater). Nonetheless, the majority of Phase III climatology and persistence ST5D forecasts had decay rates of 10-20 kt/day. Although a few STIP-predicted decay rates were as high as 30 kt/day, the majority of predictions were for decay rates of 10-20 kt/day. The JTWC intensity change forecasts during the rapid decay events were superior to all of the intensity guidance techniques.

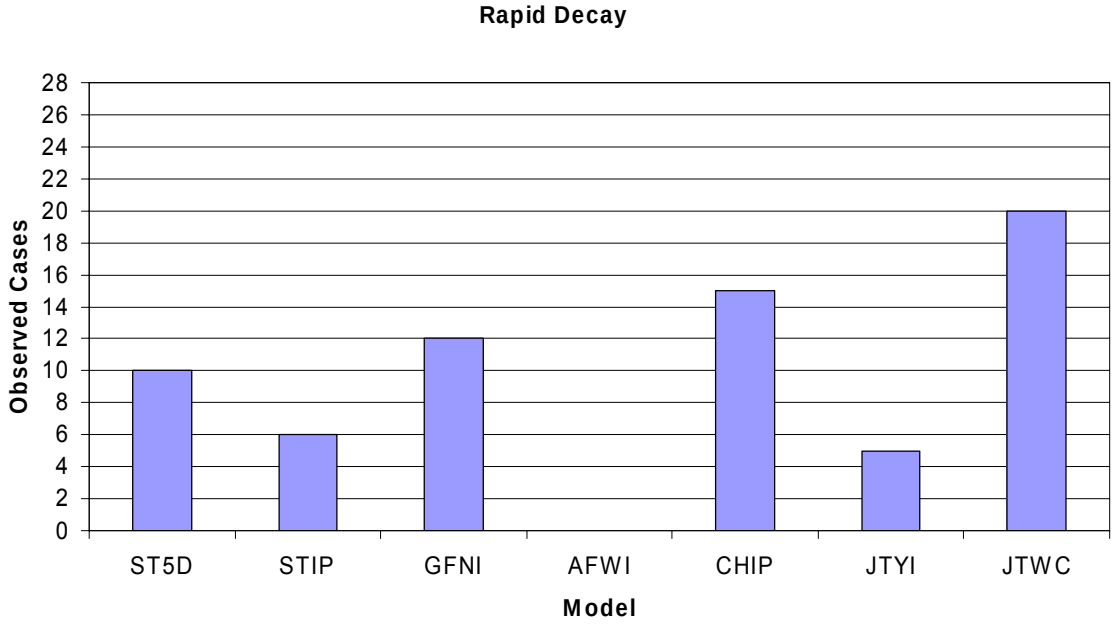

Figure 4.21 Number of cases when a forecast technique adequately predicted an observed period of rapid decay.

4. Distribution of Decay Rates During Phase III

Whereas some of the intensity guidance techniques demonstrated that they were capable of producing periods of rapid decay, an evaluation (similar to the analysis for the early intensification stage in Phase II in Chapter IV.C.4) was performed to determine quantitatively the distributions of predicted decay rates. This evaluation included all 48-h intensity change forecasts that started at the DTG of the peak intensity (or the last secondary peak) and continued into Phase III. The 48-h period was again chosen to focus on the critical decay period. The results of this evaluation are displayed in Figure 4.22 (a-g).

The actual 48-h decay rates, which are repeated in each panel of Figure 4.22, ranged from over –70 kt to only –5 kt, with the highest percentages occurring at –15, -20, -25, and –35 kt. In fact, over 60% of the decay distribution fell between –35 kt and –10 kt. Keeping in mind that this is a small sample, it is interesting that the rapid decay cases almost appear as a separate sample from the more typical decay rates. That is, the -60 to -70 kt decays in 48 h are well-separated from the primary range of decays of -15 to -35 kt. It seems likely that the very rapid decays are associated with landfall, but this needs to be verified. If the decays are due to landfall, then to correctly forecast the timing and magnitude of the decay, the techniques must have a correct track forecast. Any error in the timing of the landfall in the track forecast will necessarily introduce an error in the forecast of a rapid decay. Of course, a rapid decay of such magnitude is more likely with an intense cyclone at the initial time. The intermediate decay rates between the secondary peak and the primary decay rates may thus represent decays of a less intense cyclone.

In comparison, the ST5D has a much wider spread of 48-h intensity changes than in the observed histogram (Fig. 4.22a). Even though this technique predicted a number of rather large rates of decay, it also erroneously predicted intensification in many cases. In fact, the highest percentage of decay rates predicted by the ST5D were centered about the ‘no change’ 0 kt. Since by definition the sample was formed from the cyclones in Phase III from the last peak in the intensity curve, all values should be negative. The cluster of values

in the zero or positive range for the ST5D indicates that this climatology and persistence technique often cannot catch the beginning of the decay phase. That is, the developmental sample for the ST5D indicated that many cyclones in that location, time of year, and intensity had on average intensified rather than decayed. This false indication might arise because the decaying cases had a different direction (say moving toward land or toward a cooler ocean) than the average climatological track. That is, the forecaster should be aware that a storm with a forecast track that markedly departs from the climatological track for that location/season may not be well represented by an intensity technique based on climatology.

The intensity change distribution of the STIP (Fig. 4.22b) was similar to that of the ST5D, except this technique had an even higher percentage (~23%) of 0 kt intensity change. Over one-quarter of the STIP forecasts also erroneously predicted intensification during Phase III. Thus, the additional predictors in STIP beyond those in ST5D do not overcome the erroneous intensifications in a sample that actually only has decays. Indeed, the same intensity change dependence on location, time of year, and intensity predictors that enter into the ST5D may be exerting too much influence on the STIP prediction. Whereas the STIP predictors are calculated along the JTWC-predicted track, the anomalous track aspect should be taken into account (if the track prediction is correct).

The distribution of intensity change forecasts produced by the GFNI (Fig 4.22c) were similar to the STIP technique in that a relatively large distribution of decay rates were predicted, but also had forecasts of positive intensity changes about 20% of the time during decay. For the relatively large number of zero and positive changes, these cases indicate the factors leading to decay are not included in the model. This failure could be due to an erroneous track prediction near landfall or a failure to predict recurvature toward lower sea-surface temperatures.

Just as the AFWI tended to predict flat intensity profiles during Phase II, this technique also generated the lowest rates of decay during Phase III, with a

majority of the guidance only producing decay rates between –5 and 5 kt/48 h. This is supported by the flat intensity forecasts in the decay phase of the example in Fig. 4.20. While the AFWI forecasts tend to have small decay rates early in the forecast interval, subsequent intensification trends return the forecast to the initial value (especially by 36-48 h). Thus, the technique departs from reality by 24 h, and at later forecast intervals severely under-forecasts the rate of decay (especially during rapid decay periods).

Some very large decay rates were predicted by the CHIP model during Phase III (Fig. 4.22e). In fact, this intensity guidance technique produced some of the highest rates of decay, with several predictions of 48-h decay rates at or exceeding 50 kt. Of all the techniques, the CHIP seems to perform best in forecasting large decay rates over land and while a TC translates over lower SSTs. Nevertheless, the CHIP produced erroneous forecasts for intensification of up to 50 kt/48 h during Phase III, and the highest percentage of CHIP-predicted decay rates were centered on the 'no change' 0 kt. Such a large number of zero and positive intensity change during a decay period would cause the forecasters to ignore this technique.

The JTYI intensity change distribution (Fig. 4.22f) was also strongly peaked at 0 kt, and over 60% of the forecasts ranged between –20 kt and 0 kt. Similar to the GFNI, the JTYI is able to predict some large decay rates. As with the other intensity change techniques, the JTYI erroneously predicted some cases of intensification during the decay Phase III. Overall, the JTYI 48-h intensity change histogram most resembles the GFNI with fewer over-forecasts, but more zero values.

Lastly, over 50% of the JTWC 48-h intensity change forecasts (Fig. 4.22g) ranged between –25 kt and 0 kt, and nearly one-third of the forecasts had values at or less than –35 kt. Since the JTWC did predict rates of decay at or below –60 kt/48 h, it may indicate that forecasters favor guidance from the ST5D, STIP, or CHIP techniques (or perhaps other techniques that produced rapid decay but were not evaluated in this thesis) when anticipating high rates of decay. Perhaps

the greatest criticism of the JTWC forecasts is the too-small numbers of the primary cluster of decay rates in the range of -15 kt to -25 kt decreases over a 48-h period.

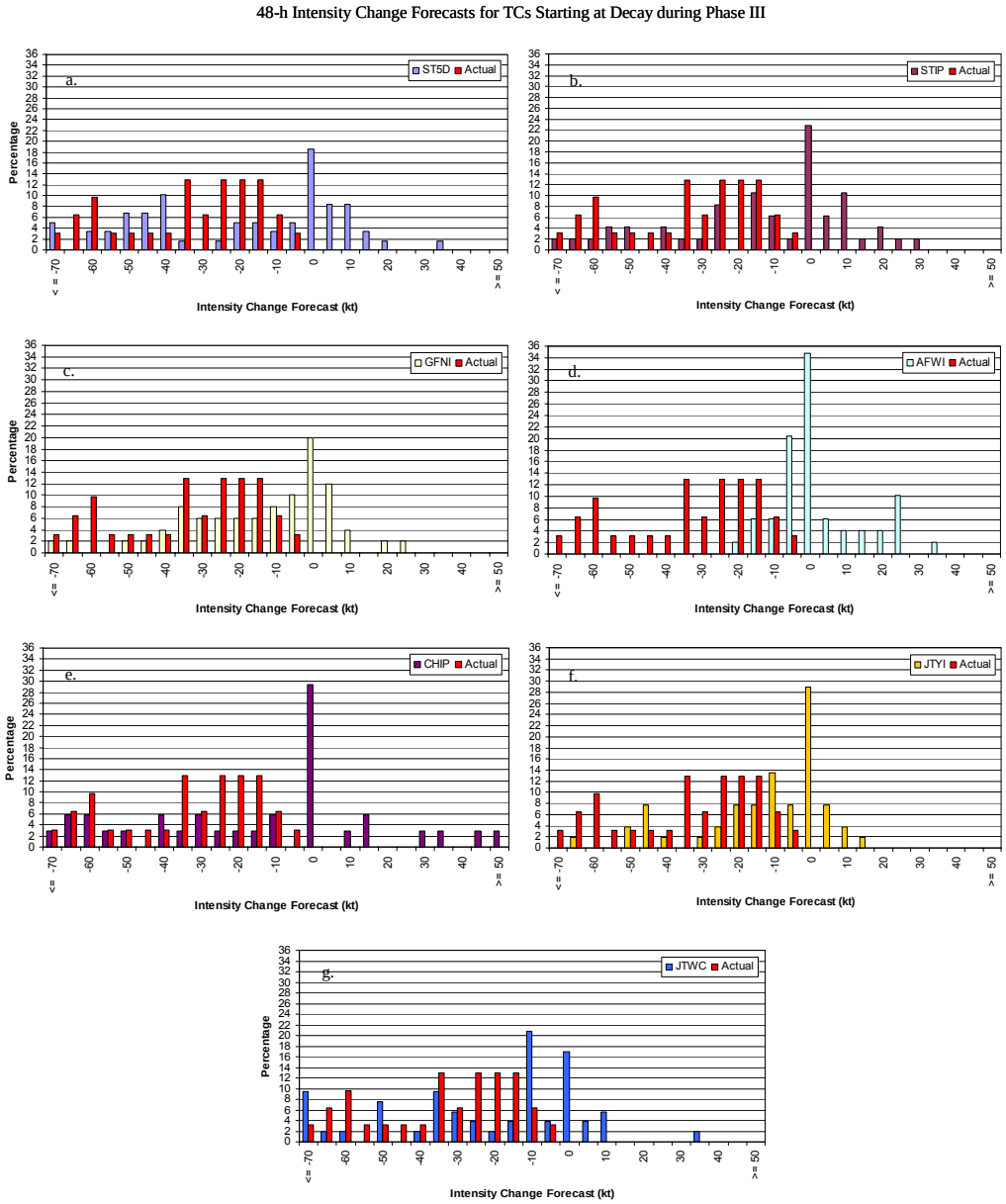

Figure 4.22 (a-g) Phase III graphs of actual 48-h intensity change distribution plotted with the 48-h intensity change distributions of all the various techniques and JTWC.

The stick diagram in Fig. 4.22h is a summary of the 48-h decay rates in Phase III similar to the summary for the intensification rates in Phase II (Fig.

4.14h). Notice that the mean observed decay is about 30 kt in 48 h with a standard deviation of almost 20 kt. However, a rather broad range of 48-h decay rates is predicted by many of the intensity guidance techniques. This is in contrast to the relatively narrow distributions of intensity change observed during Phase II (Fig 4.14h). Specifically, the ST5D, STIP, and CHIP had much larger distributions than what occurred in reality, in part due to the significant number of intensification predictions. All of the intensity guidance techniques had average rates of decay that were smaller than the actual average rate of decay. In fact, the AFWI actually had an average intensification rather than decay. Even the JTWC average decay was about 10 kt smaller than the actual decay average. If the under-forecast bias could be corrected for ST5D, STIP, and JTYI, the overall distribution would match the actual distribution.

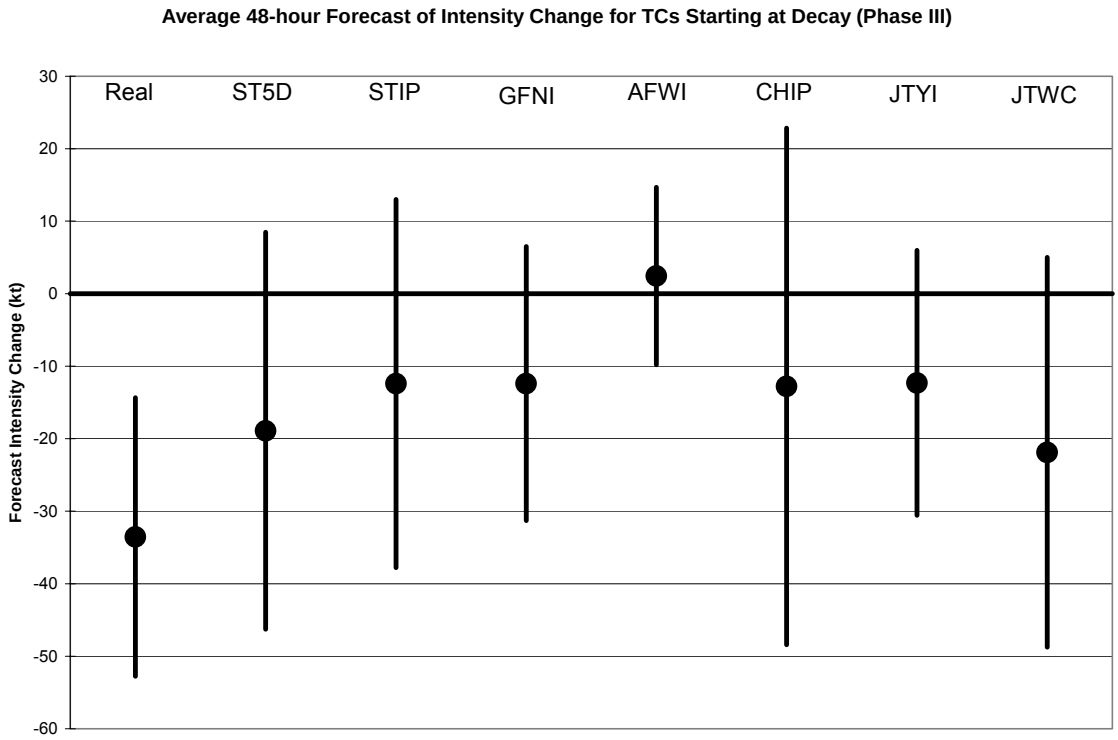

Figure 4.22h Stick diagram of 48-h intensity change distributions for TCs starting at decay. The diagram includes the actual intensity change distribution and the distributions of all the various techniques. The heavy dot represents the mean intensity change value (kt) while the length of the sticks represents plus or minus one standard deviation about the mean.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS AND RECOMMENDATIONS

A. INTENSITY GUIDANCE ACCURACY

Operational forecasters must exercise caution while evaluating the many sources of TC intensity guidance. For example, a dynamic model forecasting the most accurate track (or another model projected along the same track) is not necessarily the model with the best intensity forecast. Subjective and objective analyses support the thesis that intensity guidance in the western North Pacific does not have consistently good skill relative to a no meteorological skill technique of climatology and persistence (ST5D). A thorough intensity change forecast also requires knowledge of intensity guidance forecast traits (e.g., strengths/weaknesses and initialization processes) and a meteorological knowledge base of the synoptic/storm environment (e.g., formation regions, TC structure, and interactions between the TC and the environment). Arbitrarily selecting a particular intensity forecast versus another ignores the meteorological changes taking place within the storm and the surrounding storm environment. (Carr and Elsberry 1994).

Several objective and empirical evaluations presented in this thesis demonstrated that the available pool of statistical, statistical-dynamical, and dynamical model intensity change guidance does not skillfully predict future intensity change. On average, all of the various intensity guidance techniques and JTWC under-forecast peak intensity during Phase II. Additionally, all of the techniques and the JTWC (on average) under-forecast decay rates during Phase III (i.e., predicted decay too small). Many of the intensity guidance techniques did not capture periods of rapid intensification and decay, nor did they adequately predict intensity oscillations during Phase IIa.

Indeed, current efforts to establish a consensus approach to future intensity change are restricted, as no technique regularly outperformed guidance from the ST5D. The technique that most frequently adds skill to the ST5D is the STIP, which introduces environmental predictors such as sea-surface

temperature and vertical wind shear calculated along the JTWC-predicted track. The CHIP model also moves storms along the predicted track and predicts the SST changes under the eyewall owing to the accumulated atmospheric forcing as the front half of the storm approaches. An empirical vertical wind shear correction is also added in CHIP. The dynamic models (GFNI, AFWI, and JTYI) predict intensity as well as track, and rarely show skill relative to the ST5D.

B. INTERPRETING AVAILABLE MODEL OUTPUT

As this thesis only examined the 2003 and 2004 seasons, additional efforts will be required to assess model performance during the 2005 season and beyond. Moreover, any modifications to members of the model pool should initiate new model evaluations. With these caveats in mind, the following findings are offered as an approach for operational forecasters to interpret tropical cyclone intensity guidance based on observations and analysis from the 2003 and 2004 western North Pacific tropical seasons. A summary of these findings is presented in Table 5.1.

1. ST5D

This technique performed well for typical intensification rates (10-20 kt/day), but did not predict the onset and magnitude of rapid intensification. Forecasts beyond 48 h during the intensification stage often remained well below the verifying peak intensity value. Additionally, subjective analysis revealed that the maximum intensity forecasts of the ST5D typically remained below 110 kt (perhaps an upper bound for ST5D intensity prediction).

This technique consistently forecasts decay immediately following the first intensity peak. That is, the ST5D cannot capture decay/reintensification cycles often associated with the strongest storms. On the other hand, the ST5D can capture some rapid decay cases (e.g., up to 30 kt/day), especially for storms moving toward higher latitudes (and presumably lower SSTs and increasing vertical wind shear). However, the timing of this decay is often too early, and the ST5D missed the steepest decay rates associated with landfalling TCs. Despite the shortcomings at peak intensity and during Phase IIa, this technique

consistently provided useful intensity change guidance during Phase II and Phase III.

2. STIP

The STIP also performed well for typical intensification rates (10-20 kt/day), but even with the added dynamical component, the technique did not provide forecasts with intensification rates greater than ~20 kt/day. However, the added dynamic component may account for an increase over the ST5D in forecasts of maximum intensity. That is, despite these typical rates of intensification, the STIP technique generated maximum intensity forecasts that exceeded 110 kt.

As with the ST5D, this technique forecast persistent decay after the first intensity peak. Thus, it too could not capture the decay/reintensification cycles often associated with the strongest storms. Furthermore, the STIP did not capture periods of rapid decay, and the average decay rate was only ~10-15 kt/day instead of the real average of about 33 kt. This technique added value to the guidance produced by the ST5D in some instances, especially during intensification (Phase II).

3. GFNI

On the whole, this technique often outperformed the forecasts produced by the other dynamic, full physics techniques (i.e., AFWI and JTYI). When the GFNI correctly predicts intensification, the forecast rates are usually between 5 and 15 kt/day. However, the GFNI does not produce these intensification rates as often as the ST5D or STIP. On occasion, this technique generated intensification rates approaching 30 kt/day, but only after about 48 h into the intensification cycle. The GFNI had some success in capturing decay/reintensification cycles, as it produced the largest magnitude of intensity oscillations when compared to the other techniques. Nevertheless, the timing of these intensity oscillations was often off by 24 h or more.

Intensity guidance generated during Phase III frequently produced over-forecasts of storm intensity between 24 and 48 h (i.e., forecast guidance

remained well above the storm decay rate). However, these over-forecast tendencies diminished beyond 48 h, as the forecast decay rate approached the actual decay rate. One important consideration is that the GFNI rarely produced under-forecasts (i.e., rates of decay that exceeded the actual rate) of intensity change during Phase III.

4. AFWI

The forecasters should abandon AFWI guidance, especially if forecasting the onset of rapid intensification and decay. While a large majority of AFWI intensity change forecasts tend to undershoot actual storm intensification, there are a few cases of erroneous intensification at or just after the peak intensity has occurred—especially for intense TCs (110 kt or greater). That is, if a storm reaches an intense level, a few forecasts may continue to intensify storms beyond the PI given the location and storm environment. Conversely, the technique persistently under-forecasts observed rates of decay (especially in rapid decay cases), which results in gradual intensity declines rather than a decay.

Another problem is that the majority of AFWI intensity trends are either 'Over' or 'Under' the storm intensity, and there is less than a 50/50 chance that any 12-h forecast interval will provide 'Good' forecast trends (within 10 kt of the best track). Therefore, AFWI guidance becomes unreliable beyond 12 h, and is especially unreliable after 24 h. A possible exception to the mostly poor AFWI guidance is that the technique provided adequate intensity change forecasts for short-lived (weak) storms having only small intensity fluctuations and slow 24-h intensification rates (e.g., 5-10 kt).

5. CHIP

The CHIP technique indicated the onset of some rapid intensification/decay periods. In fact, this technique produced the highest rates of intensification/decay in relation to all of the other intensity guidance techniques. However, the timing of the intensification/decay was often off by 24 h or more. This was especially the case as the storm reached peak intensity, and the CHIP consistently forecast decay too early. Nevertheless, the CHIP-

predicted rate of decay is often closer to the actual rate of decay, at least in comparison to the other techniques. In a manner similar to the AFWI, this technique provided adequate intensity change forecasts for short-lived (weak) storms having only small intensity fluctuations (e.g., storms in Phase I of the intensity framework).

6. JTYI

The JTYI followed the AFWI technique in producing poor intensity change guidance during the intensification Phase II. This model often displayed an intensity profile similar to the AFWI model, with very modest intensification rates early (12 to 36 h), and near-constant intensity out to about 60 h. In fact, the JTYI had a strong tendency to produce under-forecasts of intensity during the formation stage (between 24 and 48 h). This may indicate that the model does not produce significant intensity change increases until the developing storm reaches a critical intensity. Thus, it is reasonable to surmise that the JTYI may provide adequate intensity change forecasts for weak storms lasting longer than two days.

This technique only produced typical rates of intensification, and these rates were not large enough to capture rapid intensification events. The flat intensity change forecasts mentioned above resulted in very large under-forecast errors at peak intensity. An evaluation of secondary peaks revealed that the JTYI is not capable of accurately predicting the timing and magnitude of intensity oscillations. This technique could not predict periods of rapid decay. Some observations indicated that the model could predict large decay rates (~30 kt/day) for landfalling TCs. While JTYI guidance is not bad at every forecast interval during Phase I, II, and III, the model rarely had skill relative to the ST5D.

24h Reasoning Indicates…	Rate (kt)	ST5D	STIP	GFNI	AFWI	CHIP	JTYI
Nearly Steady Intensity (Esp. TD in Phase I)	**0 to 5**				**X**	**X**	
Slow Intensification Rate (Esp. Moving WNW)	**5 to 10**				**X**		**X**
Average Intensification Rate (Esp. Moving WNW)	**10 to 20**	**X**	**X**				
Rapid Intensification Period (Esp. Just After Formation)	**30+**			**X**		**X**	
Intensity Greater Than 110 kt	**Any**		**X**	**X**		**X**	
Intensity Oscillations (Phase IIa)	**Any**			**X**			
Slow Decay Rate (Esp. early in Phase III)	**5 to 10**			**X**	**X**		**X**
Average Decay Rate	**10 to 20**	**X**	**X**				
Decay Past STR Axis	**Any**	**X**	**X**			**X**	
Rapid Decay at Landfall	**30+**					**X**	**X**
Rapid Decay Moving Poleward	**30+**	**X**				**X**	

Table 5.1 Summary of Phase I through Phase III model forecasts. If forecasters anticipate a certain intensification/decay rate, then the Xs in each column indicate which model is most likely to produce that forecast.

C. FUTURE WORK

Future work is needed to improve upon the problematic TC intensity forecasts in the western North Pacific (and other tropical basins). The systematic and integrated approach to tropical cyclone track forecasting detailed in Carr and Elsberry (1994) significantly improved TC track forecasting at the JTWC starting in the 1999 western North Pacific season. Perhaps a similar, repeatable method could be applied to the intensity problem. A systematic and integrated approach to TC intensity forecasting may eliminate outlying intensity forecasts, and thereby decrease the spread of model intensity forecasts about the consensus intensity forecast. For this purpose, it is necessary to examine intensity errors by: i) documenting observed intensity errors versus the official best track intensity; ii) classifying model performance against a control model representing a skill threshold; iii) inter-comparing model errors; iv) defining the synoptic/storm environment leading to a particular intensity/intensity change forecast; and v) identifying the strengths/weaknesses and sources of error for each of the techniques. Moreover, operational forecast centers should define the required accuracy for intensity forecasts—then forecasters will know what represents a ‘good’ intensity forecast at each forecast interval (e.g., 12, 24, 48, 72, 96, and 120 h).

Additionally, since the models and techniques infrequently add value relative to the ST5D, new intensity guidance techniques could be used to statistically assess when the information content of each technique is either to a larger or smaller degree of intensity change than would be expected from ST5D. That is, the objective of this approach would be to provide a combination of the guidance intelligence that will be more skillful than any of the individual techniques. When that goal is achieved, forecasters can reasonably expect to add skill over each of the individual forecasts making up the consensus.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A: WESTERN NORTH PACIFIC OPERATIONAL TROPICAL CYCLONE MODELS

The acronyms denoting various guidance techniques are explained (where known).

AFW1: Air Force MM5 (original)

AFWI: Air Force MM5 Tracker (interpolated)

BABI: Chinese Model Tracker (interpolated)

BABJ: Chinese Met Agency Warning (Beijing)

BCGZ: Chinese Met Agency Warning (from Guangzhou - not very common)

CHIP: Coupled Hurricane Intensity Prediction System

CLIM: Climatology

CLIP: Climatology and Persistence

CON_: General term for consensus model with different ending letter indicating a specific combination of models (see below)

CONG: Nonselective Consensus of one of more of the following Global Models (NGPI, JGSI, EGRI, AVNI)

CONJ: SAFA Corrected Nonselective Consensus (NGAI, EGAI, JGAI, JTYI, GFAI, COWI, AVNI, AFWI, TCLI)

CONU: As in CONJ except for NGPI, UKMI, JGSI, JTYI, GFNI, COWI, AFWI, AVNI

CONW: NGPI, EGRI, JGSI, JTYI, GFNI, COWI, AVNI, WBAI, AFWI, TCLI

COWI: COAMPS Tracker (interpolated)

DRCL: DeMaria Wind Radius CLIPER

EGAI: United Kingdom Meteorological Office (UKMO) SAFA Corrected Tracker

GFAI: GFDN Model SAFA Corrected Tracker (interpolated)

GFDL: Geophysical Fluid Dynamic Laboratory Model (original)

GFDI: Geophysical Fluid Dynamic Laboratory Model (interpolated)

GFDN: Geophysical Fluid Dynamic Laboratory—Navy Model (original)

GFNI: Geophysical Fluid Dynamic Laboratory—Navy Model (interpolated)

ICON: Intensity consensus in which forecasters can de-select any of the models before calculating average intensity change

JASD: Japan Air Self Defense Force (JASDF) Model

JAVN: AVN model with CDR Fiorino substituted tracker

JAVI: AVN model with CDR Fiorino substituted tracker (interpolated)

JGAI: JMA Spectral Model SAFA Corrected Tracker (interpolated)

JGSM: Japan Meteorological Agency (JMA) Global Spectral Model

JNGP: NOGAPS model with (CDR Fiorino) substituted tracker

JTWC: Official forecast

JTWI: Official forecast (interpolated)

JTYM: JMA Typhoon Model (regional version of JGSM)

JTYI: JMA Typhoon Model (interpolated)

JUKM: UKMO model with CDR Fiorino substituted tracker

NCON: Nonselective Consensus of NGPI, GFDI, JGSI, JTYI, and EGRI models

NGAI: NOGAPS SAFA Corrected Tracker (interpolated)

NGPS: Navy Operational Global Atmospheric Prediction System (original)

NGPI: NOGAPS Model Tracker (interpolated)

PTRO: Meteo France Model Tracker

PEST: Probabilistic Ensemble System for the Prediction of Tropical Cyclones

RECR: ‘TYAN78’ Statistical Recurver Data Set (not common anymore)

RJTD: JMA Warning

RPMM: Pagasa (Philippines) Warning

SCON: Selected Consensus of SAFA

ST5D: Statistical Typhoon Intensity Forecast (STIFOR)—5-Day

STIP: Statistical Typhoon Intensity Prediction Scheme

STID: Statistical Typhoon Intensity Prediction Scheme—Decay

STII: Statistical Typhoon Intensity Prediction Scheme (interpolated)

STRT: ‘TYAN78’ Statistical Straight Movers (not common anymore)

UKMO: United Kingdom Meteorological Office (UKMO) Model (original)

UKMI: United Kingdom Meteorological Office (UKMO) Model (interpolated)

VHHH: Hong Kong Warning

WBAR: Weber Barotropic Tropical Cyclone Track Prediction System

WBAI: Weber Barotropic Tropical Cyclone Track Prediction System (interpolated)

XTRP: Extrapolation (from past 12h intensity)

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B: STATISTICAL TYPHOON INTENSITY FORECAST FIVE-DAY MODEL (ST5D)

A. OVERVIEW

The ST5D model is a five-day TC intensity change forecast model derived from climatology and persistence. Climatology implies a long historical record of intensity changes for storms with a similar position and intensity, and persistence implies continuation of recent intensity change trends. Since forecasters typically use dynamic models to provide real-time forecast guidance, ST5D becomes a 'control' model to assess the performance (skill) of all the other models. However, ST5D may also be useful as a real-time guidance tool if only a few other techniques are available for predicting TC intensity (Knaff et al., 2003). A test of ST5D performance began during July 2001 at the Joint Typhoon Warning Center (JTWC). Initial testing showed improvement over the older three-day Statistical Typhoon Intensity Forecast (STIFOR) technique. That is, the revisions made in generating the new version of the climatology and persistence technique called ST5D also improved the performance from 12 h to 72 h as well as providing guidance out to 120 h (5 days). Chu (1994) discussed the original development of Statistical Typhoon Intensity Forecast (STIFOR) coefficients using a least-squares method for each of the seven predictors in the STIFOR model. An eighth coefficient represents the bias of the regression. The predictand in this older technique was future intensity (not intensity change) expressed as a linear combination of observable predictors.

B. MODEL DESIGN

Forecasting intensity change in ST5D requires the seven primary independent variables (predictors) listed in Table B. The dependent variable (predictand) is simply the change in intensity from an initial intensity value along the official (JTWC) forecast track. Historical storm track and intensity information used to derive the ST5D predictand was extracted from the JTWC's 1967-2000 best track datasets, including non-developing depressions that had not been

included in STIFOR. The ST5D model incorporates these non-developing depressions due to the JTWC's requirement to warn on tropical circulations exceeding 24 kt.

	Predictor	Units	Range
1	**Julian Day (JDAY):** the absolute value of the initial yearday minus yearday 248 (the climatological 'peak')	day	-248 to 117
2	**Latitude (LAT)**	deg. N	0 - 90
3	**Longitude (LON)**	deg. E	0 - 180
4	**Zonal Storm Speed (U):** E/W translation speed of storm; westerly motion is positive	kt	-50 to 50
5	**Meridional Storm Speed (V):** N/S translation speed of storm; northerly motion is positive	kt	-50 to 50
6	**Current Intensity (VMAX):** current storm intensity	kt	25 to 150
7	**Twelve-Hour Intensity Change (DVMX):** difference between current intensity and intensity 12 hours ago	kt	-100 to 100

Table B.1 Independent variables used in the ST5D.

In addition to the seven primary predictors, ST5D uses 28 secondary predictors that consist of the squares and cross-products of the original seven. This creates a potential predictor pool of 35 variables where the best predictor combinations are determined using statistical methods. By normalizing the regression coefficients, the larger magnitude coefficients exert greater influence on the forecast of the predictand (Knaff et al. 2003). Multiple linear regression techniques are then applied to develop the forecast equations from the predictor pool. Fourteen predictors are used in any ST5D intensity change forecast and three predictors are used in every forecast equation: VMAX, LAT x VMAX, and VMAX squared.

C. MODEL CHARACTERISTICS

The ST5D intensity change forecast results show a marked improvement over earlier STIFOR intensity forecasts and persistence alone. Nevertheless, caution should be exercised since this is the only year evaluated to date, and it is possible that the 2001 western North Pacific season may have had easier storms

to forecast. Future results may indeed show that the statistical formulation of ST5D is superior to the original three-day STIFOR model. In fact, the model formulation may explain critical differences between ST5D performance and its predecessor STIFOR. Knaff et al. (2003) discuss predictor selection that allows the forecast equations to actually evolve over the 120-h forecast period. The earlier climatology and persistence model included only those predictors that explained at least 0.5% of the predictand variance. Additionally, the inclusion of quadratic terms explains more variance relative to the model's linear terms.

The combined predictor JDAY x VMAX suggests intense storms—and storms that form farther from the peak of season (early season or late season)—will weaken more rapidly in the western North Pacific basin. An ST5D evaluation of intensity bias (for the 1997-2000 seasons) shows increasingly negative intensity change biases as forecast time increases (Knaff et al. 2003) A negative value indicates that, on average, western North Pacific TC intensity change is under-forecast. Maximum absolute error is smaller than the SHIFOR error at all time periods—and much smaller beyond 36 h. Moreover, the magnitude of intensity change errors levels off beyond 84 h, which indicates good model performance and forecasts better than persistence.

D. SUMMARY

While the ST5D model serves as a 'control' forecast to evaluate the skill of TC intensity forecasts, it still provides useful real-time guidance to operational forecasters because not many skillful techniques are available. Comparisons between ST5D and the older STIFOR technique demonstrate smaller mean absolute intensity errors with an overall negative intensity change bias. Since ST5D is a statistically significant improvement over STIFOR, its guidance at three, four, and five days becomes harder to outperform. In turn, this increases the challenge of operational forecasters to produce skillful TC intensity forecasts.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C: STATISTICAL TYPHOON INTENSITY PREDICTION SCHEME (STIP)

A. OVERVIEW

The STIP model is a statistical-dynamical intensity prediction model that combines the predictive strengths of statistical and dynamical approaches to intensity forecasting in the western North Pacific. The National Hurricane Center (NHC) has tested the original model, called Statistical Hurricane Intensity Prediction Scheme (SHIPS), for tropical cyclone intensity forecasts in the North Atlantic. The western North Pacific version of the model, called STIP for typhoons rather than hurricanes, began trial runs in the 2001 season at the Joint Typhoon Warning Center. Statistical contributions are extracted from variables related to climatology, persistence, and environmental factors that may influence future storm intensity. Dynamical contributions are derived from numerical weather prediction models, and include predictors related to the current and future synoptic environment. Statistical regressions of potential predictors isolate only those that have an ability to predict the dependent variable (tropical cyclone intensity changes over various intervals to 120 h). Derivation of a final predictor pool (independent variables) depends on official tropical cyclone track forecasts issued by the JTWC—several of the potential predictors are integrated along this track rather than just estimated at the initial location.

B. MODEL DESIGN

The STIP historical dataset is derived from over five years of Navy Operational Global Atmospheric Prediction System (NOGAPS) analyses of temperature, wind, water vapor pressure, and geopotential height data at many pressure levels—100, 150, 200, 250, 300, 400, 500, 700, 850, 925, and 1000 hPa. Additionally, the ocean skin temperature serves as a surrogate for sea-surface temperature. Other important historical predictor variables are gleaned from the JTWC best track data: tropical cyclone (TC) dates, positions, and intensities (to the nearest five knots) at six-hour intervals. The assumption here

is that the observed TC information and TC-related variables from the analysis fields represent long-term (i.e., climatological) synoptic conditions/regimes and TC lifecycles. These fields form the basis for the statistical portion of the model.

Since STIP is a multiple linear regression model, only the most statistically significant predictors remain in the predictor pool—based upon a combined ability to forecast intensity change. Knaff et al. (2004) describe this forecast scheme, which requires 10 predictive equations to forecast intensity change (the predictand) from an initial forecast time (in 12-hour intervals). The scheme also includes the 11 final predictors listed in Table C. These predictors are related to: i) climatology/persistence and intensity trends; ii) intensity potential as a function of current intensity and sea-surface temperature (SST); iii) the combined effect of vertical wind shear; and iv) convective instability. The statistical regression also minimizes the variance explained. Therefore, statistical techniques are more skillful for a large number of small intensity changes and are usually not as accurate for the extreme intensity changes.

	Predictor	**Forecast Hour Predictor Has Greatest Contribution to Individual Forecast Equation**
1	**DVMX:** past 12-h intensity change	**12 h**
2	**SPDX:** storm translation speed	**60 h**
3	**VMAX:** initial intensity	**12 h**
4	**VMAX²:** initial intensity squared	**12 h**
5	**MPI:** potential intensity	**24 h**
6	**MPI²:** potential intensity squared	**24 h**
7	**MPI * VMAX:** potential intensity * initial intensity	**12h**
8	**SHRD:** area-averaged (200 km to 800 km) 200-850 hPa vertical wind shear	**12 h**
9	**USHRD:** area-averaged (200 km to 800 km) 200-850 hPa zonal wind shear	**60 h**
10	**T200:** area-averaged (200 km to 800 km) temperature at 200 hPa	**36 h**
11	**RHHI:** area-averaged (200 km to 800 km) 500-300 hPa relative humidity.	**24 h**

Table C.1 Final predictors in the STIP model.

A decay-over-land version of STIP (STID) accounts for landfalling tropical cyclones and their subsequent intensity reduction. That is, the model exponentially decays TCs based upon their projected track. Tracks are interpolated to hourly positions and a land mask determines if the TC crosses a coast. This decay process begins at the landfalling intensity.

C. MODEL CHARACTERISTICS

The regression analysis revealed favorable and unfavorable conditions for TC intensification. For example, smaller values of area-averaged vertical wind shear (SHRD), and higher values of area-averaged relative humidity (RHHI), favor intensification. Conversely, higher values of SHRD and lower values of RHHI favor decay. These same findings, when applied to the predictand (DELV), reveal a general (mean) tendency for western North Pacific storms to intensify along the TC track (Knaff et al. 2004).

Knaff et al. (2004) discussed recent developmental performance of dependent data and showed that the maximum absolute error increased at all time periods. However, the trend plateaus beyond the 72-hour point: 20.7 kt (96 h); 21.2 kt (108 h); 21.8 kt (120 h). There also appears to be a preferred range of initial intensities and SSTs favoring TC intensification: SSTs greater than 28°C and initial intensities between 30 and 110 kt. Furthermore, the greatest intensity changes at 12 h and 24 h (~ 15 kt/day) occur when SSTs exceed 29.25°C and initial intensities are 75 to 80 kt.

D. SUMMARY

The STIP leverages the best statistical and dynamical tropical cyclone intensity predictors to create TC intensity guidance for use in the western North Pacific basin. The STIPS performs skillfully when the TCs intensify/decay at an average (climatological) rate. Conversely, STIPS is not as skillful for rapidly intensifying/decaying TCs. Forecast accuracy also relies on the accuracy of the forecast track and the quality of storm intensity and intensity change estimates. Additionally, the dynamical variables extracted along the forecast track require

accurate specification/initialization. A significant model assumption is the use of ocean skin temperature in lieu of ocean mixed-layer temperature. Therefore, possible (future) STIPS improvements may include the use of satellite data and ocean heat content in the predictor pool.

APPENDIX D: GEOPHYSICAL FLUID DYNAMICS LABORATORY—NAVY MODEL (GFNI) INTERPOLATED

A. OVERVIEW

The GFDN is a modified version of the Geophysical Fluid Dynamics Laboratory (GFDL) hurricane prediction system. The National Weather Service adopted this model as an 'operational' Atlantic hurricane model starting in the 1995 hurricane season. In an effort to improve tropical cyclone (TC) forecasts for the western North Pacific, researchers at the Navy's Fleet Numerical Meteorological and Oceanographic Center (FNMOC) installed the GFDL model using the Navy Operational Global Atmospheric Prediction System (NOGAPS) analyses and forecasts for the initial and boundary conditions to create what is known as GFDN. This modification provided an additional track prediction tool for JTWC forecasters, and also demonstrated improvement over purely statistical (climatological) intensity guidance beginning in the 1996 western North Pacific typhoon season (Rennick 1999). In addition to TC track forecasts, the GFDN also produces intensity change forecasts twice a day (out to 84 h) for storms in the western North Pacific.

Rennick (1999) explains the strategy for operational use of the GFDN in the western North Pacific. Each model forecast is tied to the JTWC operational procedures. Specifically, whenever one or more TCs are active within the JTWC area of responsibility, forecasters issue a TC bogus message that lists active tropical cyclone (ranked in order of precedence) locations, central pressures, radius and speed of maximum wind, and pressure of the last closed isobar.

B. MODEL DESIGN

Three computational nests are used in the GFDN: a one-degree grid to represent the synoptic environment; a 1/3-degree grid for storm environment calculations; and a 1/6-degree grid for storm inner core. The NOGAPS global analyses provide initial environmental conditions for the GFDN. As noted above,

storm-specific initialization comes from the JTWC as a TC 'bogus' message. Thereafter, NOGAPS forecast fields supply the boundary conditions for the integration on the 1/3-degree grid.

The process of removing the storm structure from the background NOGAPS flow and replacing it with an axisymmetric vortex ties the storm environmental field to the larger domain and ensures the compatibility of the specified vortex to the prediction model (Kurihara et. al. 1993). The entire vortex specification process follows:

- a crudely-resolved TC in the NOGAPS is replaced by a vortex properly specified for the GFDN model resolution;
- filtering techniques remove the vortex from the NOGAPS analysis to leave a smooth background field;
- a newly-specified vortex becomes the deviation from the background field;
- this new vortex has both axisymmetric and asymmetric components;
- the axisymmetric component results from time integration of an axisymmetric version of the model with the JTWC-specified vortex characteristics being forced ; and
- the asymmetric component from the previous model integration is inserted at the proper position.

A triply-nested, moveable-mesh system ensures that the highest resolution grid remains centered over the storm center (Kurihara et al. 1998). The vortex replacement provides excellent TC detection capability for western North Pacific tropical systems (Rennick 1999).

Several recent upgrades have been made to improve the GFDN storm intensity forecasts. The current physics include a modified cumulus parameterization scheme, enhanced surface flux calculation, an improved vertical diffusion scheme, and comprehensive radiative effects (Bender et al. 2001). Additionally, the Atlantic GFDL model was coupled with the Princeton Ocean Model (POM) in 2001, which provides a prediction of the sea-surface temperature based on the upper-ocean mixed layer physics. The POM is initialized using real-time NCEP sea-surface temperature data in conjunction with

climatological ocean data to infer sub-surface characteristics (e.g., thermocline and mixed-layer depth).

C. MODEL CHARACTERISTICS

The GFDN model is integrated with storms of tropical depression strength or greater, and the high-resolution grid over the storm center produces relatively good estimates of TC intensity. However, Rennick (1999) cites some critical model weaknesses, such as: (i) a reduced ability to predict weaker tropical systems (e.g., depressions); (ii) the timing and location of storm recurvature; and (iii) unrealistic storm intensification in strongly sheared environments. Since the GFDN is 'resource intensive,' it has a long run time—approaching 70 minutes for an 84 h forecast (GFDN web page, cited 2005).

As is shown in the text, the model had an overall negative intensity bias during 2003 and 2004, which indicates under-forecasts of intensity change. Nevertheless, the model should provide useful intensity change forecasts for storms not encountering strong vertical shear (Bender et al. 2001).

D. SUMMARY

Continued improvements to the operational GFDN have significantly increased the model's TC intensity prediction capability. Forecast studies demonstrate the good performance of the GFDN when compared to other dynamic models.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E: AIR FORCE WEATHER AGENCY MESOSCALE METEOROLOGICAL MODEL-5 (AFWI) INTERPOLATED

A. OVERVIEW

The Air Force Weather Agency (AFWA)/National Center for Atmospheric Research (NCAR)/Penn State University (PSU) Mesoscale Meteorological Model-5 (MM5) produces TC forecasts for several 'tropical' windows around the world. These tropical model forecasts are produced twice a day and provide text output of future storm location and intensity change out to 72 h. The forecast files are subsequently incorporated in the JTWC Automated Tropical Cyclone Forecast System (ATCF). Additionally, the AFWA Technology Exploitation Branch (DNXT) produces visualizations of forecast TC track (and intensity change) using special, post-processing graphics programs. These forecast graphics are located on the Joint Army Air Force Weather Information Network (JAAWIN) [Available online at https://weather.afwa.af.mil (Current as of 10 Mar 05)].

B. MODEL DESIGN

A few differences are noted between the MM5 tropical version and the operational MM5 application for midlatitude weather prediction. First, the tropical model windows are run with a different vertical resolution than midlatitude windows. There are 31 half-sigma levels in the tropical windows, versus 41 half-sigma levels in the midlatitude versions. The primary MM5 window over the western North Pacific is T19. This window contains a 15-km inner nest that is placed by the AFWA Satellite Applications Branch (XOGM). The domain is usually centered over the storm at the beginning of the model run, which may result in a fast-moving storm exiting the window before the forecast is over (private communication, 28 Oct 2004, Capt LaCroix AFWA/DNXM).

The MM5 tropical application is initialized using first-guess fields (and lateral boundary conditions) from the NCEP Global Forecast System (GFS) analyses and updated with observations (surface and/or upper air). Currently,

the operational version of the AFWI model utilizes the 3-Dimensional Variational Data Assimilation (3D-VAR) scheme to produce the initial model fields. This initialization includes a subtraction of the TC representation from the GFS first-guess fields, followed by the addition of a bogused storm vortex using the storm latitude/longitude and information on the radius of maximum winds provided by the JTWC.

Since a horizontal resolution of 45 km will lead to underestimates of storm intensity, an inner nest that has 15 km resolution will resolve more aspects of the core of the storm (Davis and Low-Nam 2001). A grid spacing less than 15 km will be required to fully resolve inner-core dynamics. The Grell cumulus parameterization technique implicitly accounts for convective processes occurring on scales much smaller than the actual grid resolution. However, Davis and Low-Nam (2001) state that the Grell scheme tends to be inactive in triggering cumulus convection, which will lead to more grid-resolved precipitation.

Future AFWA model improvements may include: (i) upgrading to the 4-Dimensional Variational Data Assimilation (4D-VAR) scheme; (ii) coupling with the Princeton Ocean Model (POM) to better simulate air/sea interaction; and (iii) switching to a moveable, fine-scale nest centered on the storm (rather than the fixed 15-km inner nest currently used) (Hausman webpage reference, cited 2005).

C. MODEL CHARACTERISTICS

The focus of the AFWA MM5 applications remains on forecasting the midlatitude mesoscale features (e.g., convective systems and frontal boundaries) that most significantly impact military operations. Since predictions for midlatitude phenomena coincide with the largest degree of military interest, more forecast verifications have been over North America and Europe. Consequently, forecasts generated in the tropical windows are usually not verified, or verified infrequently (private communication, 09 Mar 2005, Capt LaCroix AFWA/DNXM).

Therefore, there is no well-established record of AFWA typhoon forecast performance in the T19 window.

D. SUMMARY

The dynamic AFWA model with an inner grid resolution of 15 km produces TC intensity change guidance that is provided to the JTWC. While the initialization scheme and bogusing technique generally produce good initial representations of the modeled storm, other facets of model design fail to capture many aspects of TC intensity change. Although there is significant military interest in verifying midlatitude AFWA-generated forecasts, the TC forecasts produced in the MM5 tropical windows are often a lesser priority, and forecast verification statistics are lacking.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX F: COUPLED HURRICANE INTENSITY PREDICTION SYSTEM (CHIP)

A. OVERVIEW

The CHIP model is a coupled hurricane/ocean model that utilizes the recent tropical cyclone intensities in the initialization and the current storm environment to make dynamic adjustments to the initial intensity to produce an intensity forecast. Intensity forecasts are made by moving the model storm along the official forecast track from the JTWC. Emanuel et al. (2004) describe the model response to changes in the storm environment, to include the time-dependent SST under the storm center. The assumption is that internal storm fluctuations (e.g., eyewall replacement cycles) and asymmetric structures are of secondary importance to predicting the intensity. Another key assumption is that the SST modifications along the tropical cyclone track may be represented by a simple one-dimensional ocean model, rather than a fully three-dimensional ocean model.

B. MODEL DESIGN

The following environmental factors contribute to model storm intensity: pre-storm thermodynamics (i.e., potential intensity and storm-related SST anomalies); vertical shear of the horizontal wind; and dynamic upper-level perturbations (i.e., short waves). Emanuel (1995a) designed CHIP using the assumptions of environmental gradient flow and hydrostatic balance within an axisymmetric storm. The vertical structure of the vortex remains close to neutral stability, so that moist air ascending from the sea surface to some pressure level aloft follows moist adiabats (lines of constant θ_e). The SST under the center then has a crucial role because it determines the θ_e value for the ascending air parcels, and thus the potential warming of the air. Additional atmospheric considerations are given below:

- vertical structure is determined by radial distribution of planetary boundary layer moist entropy and by vorticity at the tropopause;
- moist convection is represented by one-dimensional updraft and downdraft plumes; and
- steady-state intensity is based on potential intensity (the maximum steady intensity a storm would have if the heat input is maximized via air/ocean sensible and latent heat flux).

A simple one-dimensional ocean model is used to predict SST changes along the official forecast track (Schade and Emanuel 1999). The CHIP ocean component requires an initial ocean state described by the following variables: SST; mixed layer depth; temperature jump at the base of the mixed layer; and the lapse rate below mixed layer. Conventional aerodynamic formulae are then used to describe air/sea surface fluxes and account for mixed-layer momentum processes due to surface (wind) stress driving large currents and entrainment. When the critical value of the bulk Richardson number (ratio of vertical shear of ocean current to the stability) is exceeded, mixing occurs at the base of the mixed layer, which decreases the SST and increases the mixed layer depth. The atmospheric model responds to this sea-surface temperature/upper-ocean thermal structure modification under the eye. Inertial oscillations in the mixed layer wake are not addressed in this simple, one-dimensional ocean model.

Model initialization begins with past and predicted six-hour storm positions provided by the JTWC and the intensities are linearly interpolated in time to coincide with the model's six-hour time step. The environment at storm inception is used to calculate the storm potential intensity based on the latest weekly sea-surface temperature and high-resolution (1° x 1°) atmospheric temperature analyses from 0000 UTC on the day prior to storm inception (from National Centers for Environmental Prediction—NCEP). The Center for Land-Atmosphere Prediction (COLA) utilizes an algorithm described in Bister and Emanuel (2002) to calculate storm potential intensity. Since real-time ocean state analyses are not available, monthly mean climatological data are used to define mixed layer depth and the temperature jump at the base of the mixed layer. These values

are linearly interpolated in space and time to match the storm best track or forecast position.

Dissipative effects are captured via: (i) a shear parameterization scheme and (ii) a landfall algorithm. First, vertical shear estimates originate from averaged 850 to 200-hPa NCEP horizontal wind analyses and forecast models (e.g., GFS). This averaging effectively eliminates shear associated with the storm circulation. The NCEP-derived vertical shear is used to force ventilation of the storm by trending the model equivalent potential temperature toward a climatological mid-tropospheric equivalent potential temperature. Therefore, large vertical shear values represent near-climatological equivalent potential temperature values, which are less conducive to sustaining deep convection. Knowing the prior storm intensities, the SSTs along the prior track, and using this mid-tropospheric equivalent potential temperature as a proxy for vertical shear, the model is integrated with values of this proxy. The value that leads to the best representation of the intensity history is selected and then used for launching the intensity forecast from CHIP.

Second, the Emanuel hurricane model uses a landfall algorithm to adjust intensity forecasts as the storm center comes ashore. An enthalpy exchange coefficient accounts for sensible and latent heat fluxes between land and overlying storm. This coefficient decreases linearly with land elevation and reaches zero at 40 meters. However, the rate of decrease is slowed if the storm passes over relatively flat terrain with standing water (e.g., the Florida Everglades). By quickly reducing the sensible and latent heat fluxes over the land, the equivalent potential temperature under the eyewall decreases rapidly, deep convection is eliminated, and the predicted intensity decreases.

C. MODEL CHARACTERISTICS

Since the Emanuel hurricane model is axisymmetric, the environment is quiescent. Therefore, an axisymmetric model ignores baroclinic effects on tropical cyclone intensity, such as storm interaction with midlatitude disturbances, and fluxes of mass and momentum are not evaluated across streamlines.

Consequently, the model is susceptible to large intensity errors in the presence of vertical wind shear. As indicated above, Emanuel's 'shear parameter' acts as a 'thermal' surrogate for vertical wind shear. Otherwise, the model develops (intensifies) storms as if no shear were present. While this setting of the mid-tropospheric equivalent potential temperature results in improvements over earlier CHIP versions that did not include the 'shear parameter', model solutions diverge when the model TC is moving into large shear environments.

Another potential error source is the calculation of the surface enthalpy and momentum fluxes, because these coefficients vary with the surface wind speed (Emanuel 1995b). These flux values are critical to storm intensity changes in the Emanuel model, and are difficult to measure. Emanuel found that, to a good approximation, setting the enthalpy and momentum flux coefficients to be equal and to increase linearly with gradient wind speed appears to work well in this model. However, the coefficients may not increase linearly beyond 28-30 m/s, which would affect the model intensity forecast.

Despite the simplicity of the hurricane/ocean model (relative to more complex dynamic models such as the Geophysical Fluid Dynamics Laboratory Model—GFDL), it does have some strengths. For example, experimental CHIP predictions in the Atlantic basin show some skill over purely statistical guidance (e.g., Statistical Hurricane Intensity Forecast, or SHIFOR) when the cyclone is not in large shear. The early success may be (in part) due to the accuracy of the one-dimensional ocean model, which nearly mirrors predicted SST decreases from ocean feedback of more sophisticated three-dimensional ocean models—at least under the leading edge in advance of the center. Another benefit is the model run time is only on the order of one minute on a typical personal computer.

D. SUMMARY

The dynamic hurricane/ocean CHIP model focuses on key atmospheric and oceanic contributions to tropical cyclone intensity. While the physical processes and air/sea interactions are very complex, the model addresses only

the critical processes of decreasing SST and environmental vertical shear that have significant roles in tropical cyclone intensity. Significant model assumptions include an axisymmetric cyclone, initial SSTs that are averaged over a week, a climatological upper-ocean thermal structure under the forecast storm track, and a negative intensity response to storm-induced SST decreases. Therefore, CHIP should perform well under the following conditions: in a low wind shear environment; when the forecast storm track overlies SSTs/upper thermal structure at or very near long-term averages; and if the storm does not significantly decrease SSTs along the projected track (e.g., mixing of the surface layer). Nevertheless, the accuracy of all CHIP forecasts is intimately tied to the accuracy of the official track forecast and the quality of storm intensities used for the initialization.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX G: JAPAN METEOROLOGICAL AGENCY TYPHOON MODEL (JTYI) INTERPOLATED

A. OVERVIEW

The Japan Meteorological Agency (JMA) Typhoon Model (JTYM) produces typhoon track and intensity forecasts over the western North Pacific. This model generates TC predictions four times a day (out to 84 h) for up to two TCs at a time. In addition to track and intensity predictions, the model also produces forecasts of the 30-kt wind radii and 50-kt wind radii (in four quadrants).

B. MODEL DESIGN

Currently, the Japan Global Spectral Model (JGSM) provides the initial and boundary conditions to the JTYM. The latest operational version of the GSM utilizes the 3-Dimensional Variational Data Assimilation (3D-VAR) scheme to produce the initial model fields. The JTYM is on a relocatable grid, such that the domain is relocated based on the initial storm position, and the model TC remains far removed from lateral boundaries. The horizontal grid increment is 24 km and 25 hybrid (sigma-pressure) layers are used in the vertical. A synthetic TC circulation (interpolated from the Global Spectral model analyses) is forced into the model initial fields. That is, the synthetic TC requires knowledge of the latitude/longitude, central pressure, and radius of 30 kt winds from forecaster analyses (JMA web document, cited 2005).

C. MODEL CHARACTERISTICS

The computing system improvements implemented during March 2001 reduced the previously-observed weakening bias for intense storms (JMA web document, cited 2005). As is shown in the text, this technique tends to forecast rather conservative rates of intensity change, especially during the intensification stage. Another characteristic is that the model tends to produce nearly-steady intensity change forecasts during periods of significant intensity change (similar to the AFWA technique described in Appendix E). However, a more complete

overview of model characteristics is currently not possible, as none of the web-based resources discusses model performance, and no published articles detailing the model characteristics were found.

D. SUMMARY

The JTYM is an additional TC track and intensity guidance technique available to forecasters at the JTWC. While JTYM intensity guidance output is available four times a day, the accuracy of each forecast often suffers during key stages in the storm life cycle.

APPENDIX H: DEFINITIONS

Above-average Intensification (Decay): forecast or observed intensification (decay) rates between 20 and 30 kt.

Average-lived Storm: TCs lasting from six to 12 days.

Broad Peaking: model or TC peak intensity that exceeds the usual shorter-duration intensity peak, and thus results in delayed decay.

Flat Forecast: forecast intensity remains nearly steady throughout the forecast period, despite increasing/decreasing best track intensity.

Intensity Jump: dramatic increase in model TC intensity from one 6 h forecast to the next ; a symptom of the model not being able to keep up with the actual rate of intensification (or decay).

Long-lived Storm: TCs lasting longer than 12 days.

Multiple-peaked Storm: two or more intensity peaks in the best track intensity profile that meet the definition of 'secondary peak'.

Overshooting: series of model intensity forecasts exceeding actual intensity change.

Phase Duration: duration of best track storm information in each of the three phases of the intensity framework (I, II, and III).

Primary Decay: steady decay following peak intensity, or final storm decay following any reintensification periods.

Primary Peak: largest observed intensity taken at the mid-point (date/time) of the best track intensity profile while at peak intensity.

Rapid Intensification (Decay): an increase (decrease) in observed storm intensity of 30 kt or greater in 24 h, 45 kt or greater in 36 h, or 60 kt or greater in 48 h.

Secondary Peak: a distinct reintensification of 15 kt or greater following an initial decay cycle of 15 kt or greater (e.g., eyewall replacement cycle).

Short-lived Storm: TCs lasting less than six days.

Slow Intensification (Decay): forecast or observed intensification (decay) rates less than 10 kt/day.

Storm Duration (Life Span): total duration of the best track storm information (days).

Storm Peak Span: duration of storm peak intensity (days).

Typical Intensification (Decay): forecast or observed intensification (decay) rates between 10 and 20 kt/day.

Undershooting: series of model intensity forecasts remaining below actual intensity change.

LIST OF REFERENCES

AFWA/DNXM, cited 2005: HQ AFWA's implementation of the Mesoscale Model Version 5 (MM5). [Available online at https://weather.afwa.af.mil/HOST_HOME/DNXM/ABOUTMM5/index.html (Current as of 10 Mar 05)].

Bender, M. A., I. Ginis, T. P. Marchok, and R. E. Tuleya, cited 2005: Changes to the GFDL hurricane forecast system for 2001 including implementation of the GFDL/URI hurricane-ocean coupled model. [Available online at http://www.gfdl.noaa.gov/research/weather/tpb_gfdl.html (Current as of 10 Mar 05)].

Bister, M., and K. A. Emanuel, 2002: Low frequency variability of tropical cyclone potential intensity 2. Climatology for 1982-1995. *J. Geophys. Res.,* **107,** 4621, doi:10.1029/2001JD000780.

Carr, L. E., III, and R. L. Elsberry, 1994: Systematic and integrated approach to tropical cyclone track forecasting. Part I. Approach overview and description of meteorological basis. Tech. Rep. NPS-MR-94-002, Naval Postgraduate School, Monterey, CA 93943-5114, 273 pp.

Chu, J. H., 1994: A regression model for the western North Pacific tropical cyclone intensity forecasts. NRL Memo. Rep. 7541-94-7215, Naval Research Laboratory, Monterey, CA 93943-5502, 33 pp.

Davis, C.A., and S. Low-Nam, 2001: The NCAR-AFWA tropical cyclone bogussing scheme. Tech. Memo to the Air Force Weather Agency (AFWA), Offutt AFB, NE, 13 pp.

DeMaria, M., and J. Kaplan, 1999: An updated hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific. *Wea. Forecasting,* **14,** 326-337.

Emanuel, K. A./MIT, cited 2005: Limits on hurricane intensity. [Available online at http://wind.mit.edu/~emanuel/holem/holem.html (current as of 04 Mar 05)].

Emanuel, K. A., 1986: An air-sea interaction theory for tropical cyclones. Part I: Steady-state maintenance. *J. Atmos. Sci.,* **43,** 585-604.

Emanuel, K. A., 1995a: The behavior of a simple hurricane model using a convective scheme based on subcloud-layer entropy equilibrium. *J. Atmos. Sci.,* **52,** 3959-3968.

Emanuel, K. A., 1995b: Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady-state model incorporating eye dynamics. *J. Atmos. Sci.,* **52,** 3969-3976.

Emanuel, K. A., C. DesAutels, C. Holloway, and R. Korty, 2004: Environmental control of tropical cyclone intensity. *J. Atmos. Sci.,* **61,** 843-858.

GFDN/FNMOC, cited 2005: Geophysical Fluid Dynamics Laboratory-Navy model (GFDN) description. [Available online at https://www.fnmoc.navy.mil/PUBLIC/MODEL_REPORTS/MODEL_SPEC/gfdn.html (Current as of 13 Mar 05)].

Harr, P. A., and R. L. Elsberry, 2000: Extratropical transition of tropical cyclones over the western North Pacific. Part I: Evolution of structural characteristics during the transition process. *Mon. Wea. Rev.,* **128,** 2613-2633.

Harr, P. A., R. L. Elsberry, and T. F. Hogan, 2000: Extratropical transition of tropical cyclones over the western North Pacific. Part II: The impact of midlatitude circulation characteristics. *Mon. Wea. Rev.,* **128,** 2634-2653.

Holland, G. J., 1997: The maximum potential intensity of tropical cyclones. *J. Atmos. Sci.,* **54,** 2519-2541.

Holliday, C. R., and A. H. Thompson, 1979: Climatological characteristics of rapidly intensifying typhoons. *Mon. Wea. Rev.,* **107,** 1022-1034.

Hausman, S./AFWA, cited 2005: Models, models everywhere…literally. [Available online at https://afweather.afwa.af.mil/observer/JUL_AUG_2001/Models,_Models_everywhere...literally (Current as of 10 Mar 2005)].

Jeffries, R. A., and E. M. Fukada/JTWC, cited 2005: Consensus approach to TC forecasting. [Available online at http:www.npmoc.navy.mil/jtwc/atcr/2002atdr/ch6/chap6_page1.html (Current as of 03 Mar 05)].

JMA, cited 2005: Numerical model descriptions [Available online at http://www.jma.go.jp/JMA_HP/jma/jma-eng/jma-center/nwp/outline-nwp/pdf/ol4_6.pdf (Current as of 10 Mar 05)].

Klein, P. M., P. A. Harr, and R. L. Elsberry, 2000: Extratropical transition of western North Pacific tropical cyclones: An overview and conceptual model of the transformation stage. *Wea. Forecasting,* **15,** 373-395.

Knaff, J. A., M. DeMaria, C. R. Sampson, and J. M. Gross, 2003: Statistical, five-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting,* **18,** 80-92.

Knaff, J. A., M. DeMaria, and C. R. Sampson, 2004: An operational statistical typhoon intensity prediction scheme for the western North Pacific. Manuscript submitted to *Wea. Forecasting*.

Miller, B. I., 1958: On the maximum intensity of hurricanes. *J. Meteor.,* **15,** 184-195.

Rennick, M. A., 1999: Performance of the Navy's tropical cyclone prediction model in the western North Pacific basin during 1996. *Wea. Forecasting,* **14,** 297-305.

Schade, L. R., and K. A. Emanuel, 1999: The ocean's effect on the intensity of tropical cyclones: Results from a simple coupled atmosphere-ocean model. *J. Atmos. Sci.,* **56,** 642-651.

Willoughby, H. E., J. A. Clos, and M. G. Shoreibah, 1982: Concentric eye walls, secondary wind maxima, and the evolution of the hurricane vortex. *J. Atmos. Sci.,* **39,** 395-411.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Air Force Weather Technical Library
   Asheville, North Carolina

4. Air Force Institute of Technology
   Wright-Patterson Air Force Base, Ohio

5. Professor Philip A. Durkee
   Naval Postgraduate School
   Monterey, California

6. Professor Russell L. Elsberry
   Naval Postgraduate School
   Monterey, California

7. Mark A. Boothe
   Naval Postgraduate School
   Monterey, California

8. Director, Joint Typhoon Warning Center
   Naval Pacific Meteorology and Oceanography Center
   Pearl Harbor, Hawaii

9. Captain Jason S. Blackerby
   Joint Typhoon Warning Center
   Pearl Harbor, Hawaii